Sequence Alignments

```
                        GenCore version 5.1.6
               Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         April 22, 2004, 12:46:56 ; Search time 48 Seconds
                                           (without alignments)
                                           1572.453 Million cell updates/sec

Title:          US-09-978-191A-506
Perfect score:  1505
Sequence:       1 MRGSQEVLLMWLLVLAVGGT..........SEQISFLEEQLGSCSCKKDS 273

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1133595 seqs, 276475211 residues

Total number of hits satisfying chosen parameters:      1133595

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
               1:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
               2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
               3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
               4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
               5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
               6:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
               7:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
               8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
               9:  /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
              10:  /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
              11:  /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
              12:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
              13:  /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
              14:  /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
              15:  /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
              16:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
              17:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
              18:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                 SUMMARIES
                     %
Result              Query
  No.     Score     Match Length DB  ID                        Description
-----------------------------------------------------------------------------
    1     1505      100.0    273  9  US-09-790-264-10          Sequence 10, Appl
    2     1505      100.0    273  9  US-09-978-295A-506        Sequence 506, App
    3     1505      100.0    273  9  US-09-978-295A-508        Sequence 508, App
    4     1505      100.0    273  9  US-09-978-697-506         Sequence 506, App
    5     1505      100.0    273  9  US-09-978-697-508         Sequence 508, App
    6     1505      100.0    273  9  US-09-978-192A-506        Sequence 506, App
    7     1505      100.0    273  9  US-09-978-192A-508        Sequence 508, App
    8     1505      100.0    273  9  US-09-999-832A-506        Sequence 506, App
    9     1505      100.0    273  9  US-09-999-832A-508        Sequence 508, App
   10     1505      100.0    273 10  US-09-978-189-506         Sequence 506, App
   11     1505      100.0    273 10  US-09-978-189-508         Sequence 508, App
   12     1505      100.0    273 10  US-09-978-608A-506        Sequence 506, App
   13     1505      100.0    273 10  US-09-978-608A-508        Sequence 508, App
   14     1505      100.0    273 10  US-09-978-585A-506        Sequence 506, App
   15     1505      100.0    273 10  US-09-978-585A-508        Sequence 508, App
   16     1505      100.0    273 10  US-09-978-191A-506        Sequence 506, App
   17     1505      100.0    273 10  US-09-978-191A-508        Sequence 508, App
   18     1505      100.0    273 10  US-09-978-403A-506        Sequence 506, App
   19     1505      100.0    273 10  US-09-978-403A-508        Sequence 508, App
   20     1505      100.0    273 10  US-09-978-564A-506        Sequence 506, App
   21     1505      100.0    273 10  US-09-978-564A-508        Sequence 508, App
   22     1505      100.0    273 10  US-09-999-833A-506        Sequence 506, App
   23     1505      100.0    273 10  US-09-999-833A-508        Sequence 508, App
   24     1505      100.0    273 10  US-09-981-915A-506        Sequence 506, App
   25     1505      100.0    273 10  US-09-981-915A-508        Sequence 508, App
   26     1505      100.0    273 10  US-09-978-824-506         Sequence 506, App
   27     1505      100.0    273 10  US-09-978-824-508         Sequence 508, App
   28     1505      100.0    273 10  US-09-918-585A-506        Sequence 506, App
   29     1505      100.0    273 10  US-09-918-585A-508        Sequence 508, App
   30     1505      100.0    273 10  US-09-978-423A-506        Sequence 506, App
   31     1505      100.0    273 10  US-09-978-423A-508        Sequence 508, App
   32     1505      100.0    273 10  US-09-978-193A-506        Sequence 506, App
   33     1505      100.0    273 10  US-09-978-193A-508        Sequence 508, App
   34     1505      100.0    273 10  US-09-999-830A-506        Sequence 506, App
   35     1505      100.0    273 10  US-09-999-830A-508        Sequence 508, App
   36     1505      100.0    273 10  US-09-978-757A-506        Sequence 506, App
   37     1505      100.0    273 10  US-09-978-757A-508        Sequence 508, App
   38     1505      100.0    273 10  US-09-978-187B-506        Sequence 506, App
   39     1505      100.0    273 10  US-09-978-187B-508        Sequence 508, App
   40     1505      100.0    273 10  US-09-978-643A-506        Sequence 506, App
   41     1505      100.0    273 10  US-09-978-643A-508        Sequence 508, App
   42     1505      100.0    273 10  US-09-978-375A-506        Sequence 506, App
   43     1505      100.0    273 10  US-09-978-375A-508        Sequence 508, App
   44     1505      100.0    273 10  US-09-978-298A-506        Sequence 506, App
   45     1505      100.0    273 10  US-09-978-298A-508        Sequence 508, App

                                ALIGNMENTS

RESULT 1
US-09-790-264-10
; Sequence 10, Application US/09790264
; Patent No. US20020028508A1
; GENERAL INFORMATION:
;  APPLICANT: Holtzman, Douglas A.
;  APPLICANT:  Goodearl, Andrew D.J.
;  APPLICANT:  McCarthy, Sean A.
;  TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING
;  TITLE OF INVENTION:  PROGNASTIC, DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER
;  TITLE OF INVENTION:  USES
;  FILE REFERENCE: 07334-322001
;  CURRENT APPLICATION NUMBER: US/09/790,264
;  CURRENT FILING DATE:  2001-02-21
;  PRIOR APPLICATION NUMBER: US 09/065,661
;  PRIOR FILING DATE: 1998-04-23
;  PRIOR APPLICATION NUMBER: US 09/298,531
;  PRIOR FILING DATE: 1999-04-23
;  PRIOR APPLICATION NUMBER: US 09/065,363
;  PRIOR FILING DATE: 1998-04-23
;  PRIOR APPLICATION NUMBER: US 09/337,930
;  PRIOR FILING DATE: 1999-06-22
;  PRIOR APPLICATION NUMBER: US 09/102,705
;  PRIOR FILING DATE: 1998-06-22
;  PRIOR APPLICATION NUMBER: US 09/363,630
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 09/124,538
;  PRIOR FILING DATE: 1998-07-29
;  NUMBER OF SEQ ID NOS: 68
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 273
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SIGNAL
;   LOCATION: (1)...(22)
US-09-790-264-10
```

```
  Query Match             100.0%;  Score 1505;  DB 9;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-113;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273

RESULT 2
US-09-978-295A-506
; Sequence 506, Application US/09978295A
; Patent No. US20020156006A1
; GENERAL INFORMATION:
;  APPLICANT: Ashkenazi, Avi
;  APPLICANT:  Baker Kevin P.
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan
;  APPLICANT:  Ferrara, Napoleon
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, Audrey
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, J. Christopher
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth J
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Kuo, Sophia S.
;  APPLICANT:  Napier, Mary A.
;  APPLICANT:  Pan, James;
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Shelton, David L.
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: P2630P1C11
;  CURRENT APPLICATION NUMBER: US/09/978,295A
;  CURRENT FILING DATE:  2001-10-15
;  PRIOR APPLICATION NUMBER: 09/918585
;  PRIOR FILING DATE: 2001-07-30
;  PRIOR APPLICATION NUMBER: 60/062250
;  PRIOR FILING DATE: 1997-10-17
;  PRIOR APPLICATION NUMBER: 60/064249
;  PRIOR FILING DATE: 1997-11-03
;  PRIOR APPLICATION NUMBER: 60/065311
;  PRIOR FILING DATE: 1997-11-13
;  PRIOR APPLICATION NUMBER: 60/066364
;  PRIOR FILING DATE: 1997-11-21
;  PRIOR APPLICATION NUMBER: 60/077450
;  PRIOR FILING DATE: 1998-03-10
;  PRIOR APPLICATION NUMBER: 60/077632
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077641
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077649
;  PRIOR FILING DATE: 1998-03-11
;  PRIOR APPLICATION NUMBER: 60/077791
;  PRIOR FILING DATE: 1998-03-12
;  PRIOR APPLICATION NUMBER: 60/078004
;  PRIOR FILING DATE: 1998-03-13
;  PRIOR APPLICATION NUMBER: 60/078886
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078936
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078910
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/078939
;  PRIOR FILING DATE: 1998-03-20
;  PRIOR APPLICATION NUMBER: 60/079294
;  PRIOR FILING DATE: 1998-03-25
;  PRIOR APPLICATION NUMBER: 60/079656
;  PRIOR FILING DATE: 1998-03-26
;  PRIOR APPLICATION NUMBER: 60/079664
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079689
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079663
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079728
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079786
;  PRIOR FILING DATE: 1998-03-27
;  PRIOR APPLICATION NUMBER: 60/079920
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: 60/079923
;  PRIOR FILING DATE: 1998-03-30
;  PRIOR APPLICATION NUMBER: 60/080105
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080107
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080165
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080194
;  PRIOR FILING DATE: 1998-03-31
;  PRIOR APPLICATION NUMBER: 60/080327
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080328
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080333
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/080334
;  PRIOR FILING DATE: 1998-04-01
;  PRIOR APPLICATION NUMBER: 60/081070
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081049
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081071
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081195
;  PRIOR FILING DATE: 1998-04-08
;  PRIOR APPLICATION NUMBER: 60/081203
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: 60/081229
;  PRIOR FILING DATE: 1998-04-09
;  PRIOR APPLICATION NUMBER: 60/081955
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081817
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081819
;  PRIOR FILING DATE: 1998-04-15
;  PRIOR APPLICATION NUMBER: 60/081952
```

```
F;702-704/Region: cell attachment (R-G-D) motif
F;714-750/Domain: EGF homology <EG3>
F;762-800/Domain: EGF homology <EG4>
F;806-839/Domain: EGF homology <EG5>
F;849-919/Domain: thyroglobulin type I repeat homology <THY1>
F;990-1032/Domain: LDL receptor YWTD-containing repeat homology <YW1>
F;1033-1075/Domain: LDL receptor YWTD-containing repeat homology <YW2>
F;1076-1120/Domain: LDL receptor YWTD-containing repeat homology <YW3>
F;1121-1160/Domain: LDL receptor YWTD-containing repeat homology <YW4>
F;1161-1197/Domain: LDL receptor YWTD-containing repeat homology <YW5>
F;1212-1243/Domain: EGF homology <EG6>
F;289,296/Binding site: sulfate (Tyr) (covalent) #status predicted
F;729,819/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F;756/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of laminin) #status predi
F;1137/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             11.5%;  Score 173.5;  DB 1;  Length 1247;
  Best Local Similarity   25.8%;  Pred. No. 4.6e-05;
  Matches   80;  Conservative   23;  Mismatches   98;  Indels  109;  Gaps   18;

Qy           20 TEHAYRPGRRV-----CAVRAHG--------DPVSESFVQRVYQPFLTTCDGHRACSTYR 66
                |  | ||| |      |::   |        |  ||       ||  : |  |  |: :
Db          681 TNAACRPGPRTQFTCECSIGFRGDGRTCYDIDECSE-------QP--SVCGSHTICNNH- 730

Qy           67 TIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPP---CRNG--------- 114
                              |   |  |  |:: :    | | | : | |   |  |
Db          731 --------------PGTFRCECVEGYQFSD--EGTCVAVVDQRPINYCETGLHNCDIPQR 774

Qy          115 GSCVQPG----RCRCPAGWRGD--TCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHS 168
                  |:  |     | |  |: ||   || |||||   |      | || ||: |||  |:
Db          775 AQCIYTGGSSYTCSCLPGFSGDGQACQ-DVDECQPSRCHPDAFCYNTPGSFTCQCKPGY- 832

Qy          169 LSADGTLCVP------------------KGGPPRVAPNPTG------------------- 191
                   ||  |||                   |      | | |
Db          833 -QGDGFRCVPGEVEKTRCQHEREHILGAAGATDPQRPIPPGLFVPECDAHGHYAPTQCHG 891

Qy          192 -------VDSAMKEEVQRLQSRVDLLEEKLQLVLAPLH-SLASQALEHGLPDPGSLLVHS 243
                       ||   : ||:  ::|  :    |  |  |:|   |       || ||: |:
Db          892 STGYCWCVDRDGR-EVEGTRTRPGMTPPCLSTVAPPIHQGPAVPTAVIPLP-PGTHLL-- 947

Qy          244 FQQLGRIDSL 253
                | | |:|: |
Db          948 FAQTGKIERL 957

RESULT 15
A24420
notch protein - fruit fly (Drosophila melanogaster)
N;Alternate names: neurogenic repetitive locus protein
C;Species: Drosophila melanogaster
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A24420; A24768; S09358; A05267
R;Kidd, S.; Kelley, M.R.; Young, M.W.
Mol. Cell. Biol. 6, 3094-3108, 1986
A;Reference number: A24420; MUID:87064624; PMID:3097517
A;Accession: A24420
A;Molecule type: DNA
A;Residues: 1-2703 <KID>
A;Cross-references: GB:K03508; NID:g157991; PIDN:AAA28725.1; PID:g157993
R;Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
Cell 43, 567-581, 1985
A;Reference number: A24768; MUID:86079539; PMID:3935325
A;Accession: A24768
A;Molecule type: mRNA
A;Residues: 1-48,'I',50-118,'R',120-230,'I',232-256,'N',258-266,'A',268-872,'R',874-958,
A;Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2044
R;Tautz, D.
Nucleic Acids Res. 17, 6463-6471, 1989
A;Title: Hypervariability of simple sequences as a general source for polymorphic DNA ma
A;Reference number: S09358; MUID:89385974; PMID:2780284
A;Accession: S09358
A;Molecule type: DNA
A;Residues: 2505-2551,'QQQQ',2552-2576,'E',2578-2604 <TAU>
```

```
R;Wharton, K.A.; Yedvobnick, B.; Finnerty, V.G.; Artavanis-Tsakonas, S.
Cell 40, 55-62, 1985
A;Title: opa: a novel family of transcribed repeats shared by the Notch locus and othe
A;Reference number: A05267; MUID:85099329; PMID:2981631
A;Accession: A05267
A;Molecule type: DNA
A;Residues: 2504-2576,'E',2578-2611 <WHA2>
C;Genetics:
A;Gene: notch; opa
A;Cross-references: FlyBase:FBgn0004647
A;Map position: 8.96-9.36
A;Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: differentiation; tandem repeat; transmembrane protein
F;27-43/Domain: transmembrane #status predicted <TMM1>
F;297-328/Domain: EGF homology <EGX1>
F;530-561/Domain: EGF homology <EGF1>
F;568-599/Domain: EGF homology <EGF>
F;988-1019/Domain: EGF homology <EGX2>
F;1064-1095/Domain: EGF homology <EGF3>
F;1187-1218/Domain: EGF homology <EGX3>
F;1746-1762/Domain: transmembrane #status predicted <TMM2>
F;1950-1982/Domain: ankyrin repeat homology <AN1>
F;1983-2015/Domain: ankyrin repeat homology <AN2>
F;1988-2004/Domain: transmembrane #status predicted <TMM3>
F;2017-2049/Domain: ankyrin repeat homology <AN3>
F;2050-2082/Domain: ankyrin repeat homology <AN4>
F;2083-2115/Domain: ankyrin repeat homology <AN5>
F;2538-2568/Region: glutamine-rich
F;2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

  Query Match             11.5%;  Score 173;  DB 1;  Length 2703;
  Best Local Similarity   26.7%;  Pred. No. 9.9e-05;
  Matches   59;  Conservative   19;  Mismatches   67;  Indels   76;  Gaps   13;

Qy           18 GGTEHAYRPGRRVCAVRAH--GDPVSESFVQRVYQPFLTTCDGHRACSTYRTIYRTAYRR 75
                |||      |:  ||  :|  ||          |      |:  | |    |   |
Db           70 GGTCVTQLNGKTYCACDSHYVGD----------YCEHRNPCNSMR-CQNGGTCQVTFRNG 118

Qy           76 SPGLAPARPRYAC-CP-GWKRT---SGLPGACGAAIC----------------------- 107
                 ||:       :| || |:  :     :| ||    |
Db          119 HPGI-------SCKCPLGFDESLCEIAVPNACDHVTCLNGGTCQLKTLEEYTCACANGYT 171

Qy          108 -----------QPPCRNGGSCV-----QPGRCRCPAGWRGDTCQSDVDECSA---RRGGC 148
                             ||||| :|          | || |: ||||  |::|| :   : ||
Db          172 GERCETKNLCASSPCRNGATCTALAGSSSFTCSCPPGFTGDTCSYDIEECQSNPCKYGG- 230

Qy          149 PQRCINTAGSYWCQCWEGHSLSADGTLCVPKGGPPRVAPNP 189
                   |:|| ||| | |  |::    |  |  |  |   :|:|
Db          231 --TCVNTHGSYQCMCPTGYT----GKDCDTKYKP--CSPSP 263

Search completed: April 22, 2004, 12:47:22
Job time : 22 secs
```

```
S78549
notch3 protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 08-Sep-2002
C;Accession: S78549; S71825
R;Joutel, A.; Tournier-Lasserve, E.
submitted to the EMBL Data Library, April 1997
A;Reference number: S78549
A;Accession: S78549
A;Molecule type: mRNA
A;Residues: 1-2321 <JOU1>
A;Cross-references: EMBL:U97669; NID:g2668591; PIDN:AAB91371.1; PID:g2668592
R;Joutel, A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chabriat, H.; Mouton, P.; Alamowitc
x, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserve, E.
Nature 383, 707-710, 1996
A;Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke
A;Reference number: S71825; MUID:97032728; PMID:8878478
A;Accession: S71825
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2
A;Cross-references: EMBL:U97669
C;Genetics:
A;Gene: notch3
A;Map position: 19p13.1
C;Function:
A;Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: tandem repeat; transmembrane protein
F;123-155/Domain: EGF homology <EGX1>
F;162-194/Domain: EGF homology <EGF1>
F;240-271/Domain: EGF homology <EGX2>
F;318-349/Domain: EGF homology <EGF>
F;473-504/Domain: EGF homology <EGX3>
F;853-884/Domain: EGF homology <EGF3>
F;928-959/Domain: EGF homology <EGX4>
F;1838-1870/Domain: ankyrin repeat homology <AN1>
F;1871-1903/Domain: ankyrin repeat homology <AN2>
F;1905-1937/Domain: ankyrin repeat homology <AN3>
F;1938-1970/Domain: ankyrin repeat homology <AN4>
F;1971-2003/Domain: ankyrin repeat homology <AN5>

  Query Match             11.7%;  Score 175.5;  DB 2;  Length 2321;
  Best Local Similarity   34.5%;  Pred. No. 5.7e-05;
  Matches   50;  Conservative   14;  Mismatches   44;  Indels   37;  Gaps   11;

Qy          56 CDGHRACSTYRTIYRTAYRRSPGLAPARPRYAC-CPGWKR--TSGLPGACGAAICQPPCR 112
               | |   | :              :     |::| ||   |     ||  |       ||
Db          87 CAGRGVCQS-------------SVVAGTARFSCRCPRGFRGPDCSLPDPC----LSSPCA 129

Qy         113 NGGSC-VQP-GR--CRCPAGWRGDTCQSDVDEC----SARRGGCPQRCINTAGSYWCQCW 164
               :|  | | | ||  | || |::| :|:||||||      | ||    |:|| ||: |||
Db         130 HGARCSVGPDGRFLCSCPPGYQGRSCRSDVDECRVGEPCRHGG---TCLNTPGSFRCQCP 186

Qy         165 EGHSLSADGTLCVPKGGPPRVAPNP 189
                |::    | ||     |   ||:|
Db         187 AGYT----GPLCENPAVP--CAPSP 205

RESULT 13
JP0076
nel protein - chicken
C;Species: Gallus gallus (chicken)
C;Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 13-Aug-1999
C;Accession: A38963; JP0076
R;Matsuhashi, S.; Noji, S.; Koyama, E.; Myokai, F.; Ohuchi, H.; Taniguchi, S.; Hori, K.
Dev. Dyn. 203, 212-222, 1995
A;Title: New gene, nel, encoding a Mr 93K protein with EGF-like repeats is strongly expr
A;Reference number: A38963; MUID:95383734; PMID:7655083
A;Accession: A38963
A;Molecule type: mRNA
A;Residues: 1-835 <MAT>
A;Cross-references: DDBJ:D45365
```

```
A;Experimental source: 9-day embryo
R;Matsuhashi, S.; Noji, S.; Koyama, E.; Myokai, F.; Ohuchi, H.; Taniguchi, S.; Hori, K
submitted to JIPID, January 1995
A;Description: A new gene, nel, encoding a Mr 93K protein with EGF-like repeats is str
A;Reference number: JP0076
A;Accession: JP0076
A;Molecule type: mRNA
A;Residues: 1-835 <MA2>
A;Cross-references: DDBJ:D45365
A;Experimental source: 9-day embryo
C;Superfamily: von Willebrand factor type C repeat homology; EGF homology
F;273-333/Domain: von Willebrand factor type C repeat homology <VWC>
F;395-592/Region: EGF-like repeats
F;444-480/Domain: EGF homology <EGF1>
F;486-521/Domain: EGF homology <EGF>
F;525-552/Domain: EGF homology <EGF2>

  Query Match             11.6%;  Score 174;  DB 2;  Length 835;
  Best Local Similarity   42.9%;  Pred. No. 3e-05;
  Matches   30;  Conservative   11;  Mismatches   27;  Indels    2;  Gaps    1;

Qy         100 GACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQR--CINTAG 157
               |    | |:  |||||:|:    | || |: | :|::|:||||     |  |  |||  |
Db         518 GTVCKAFCKDGCRNGGACIASNVCACPQGFTGPSCETDIDECSDGFVQCDSRANCINLPG 577

Qy         158 SYWCQCWEGH 167
                | |:| :|:
Db         578 WYHCECRDGY 587

RESULT 14
MMHUND
nidogen precursor - human
N;Alternate names: entactin
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Oct-2000
C;Accession: A33322; A32437; A61367
R;Nagayoshi, T.; Sanborn, D.; Hickok, N.J.; Olsen, D.R.; Fazio, M.J.; Chu, M.L.; Knowlt
DNA 8, 581-594, 1989
A;Title: Human nidogen: complete amino acid sequence and structural domains deduced fr
A;Reference number: A33322; MUID:90091745; PMID:2574658
A;Accession: A33322
A;Molecule type: mRNA
A;Residues: 1-1247 <NAG>
A;Cross-references: EMBL:M30269
R;Olsen, D.R.; Nagayoshi, T.; Fazio, M.; Mattei, M.G.; Passage, E.; Weil, D.; Timpl, R
Am. J. Hum. Genet. 44, 876-885, 1989
A;Title: Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to
A;Reference number: A32437; MUID:89270475; PMID:2471408
A;Accession: A32437
A;Molecule type: mRNA
A;Residues: 667-1247 <OLS>
A;Cross-references: EMBL:M27445; NID:g602466; PIDN:AAA57261.1; PID:g602467
A;Note: the authors translated the codon AAG for residue 966 as Cys
R;Fazio, M.J.; O'Leary, J.; Kaehaeri, V.M.; Chen, Y.Q.; Saitta, B.; Uitto, J.
J. Invest. Dermatol. 97, 281-285, 1991
A;Title: Human nidogen gene: structural and functional characterization of the 5'-flan
A;Reference number: A61367; MUID:91302882; PMID:1906509
A;Accession: A61367
A;Molecule type: DNA
A;Residues: 1-28 <FAZ>
C;Comment: This protein is a basement membrane glycoprotein that forms a complex with
C;Genetics:
A;Gene: GDB:NID
A;Cross-references: GDB:120236; OMIM:131390
A;Map position: 1q43-1q43
C;Superfamily: nidogen; EGF homology; LDL receptor YWTD-containing repeat homology; th
C;Keywords: basement membrane; beta-hydroxyasparagine; calcium binding; cell binding;
protein
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1247/Product: nidogen #status predicted <MAT>
F;390-425/Domain: EGF homology <EG1>
F;672-708/Domain: EGF homology <EG2>
```

```
Db         504 CQPGYVGN----GTICKAFCEEGCRYGGTCVAPNKCVCPSGFTGSHCEKDIDECAEGFVE 559

Qy         148 CP--QRCINTAGSYWCQCWEGH----SLSADGTLCV 177
               |    ||:|  | | |:|  |     : |  |  |:
Db         560 CHNYSRCVNLPGWYHCECRSGFHDDGTYSLSGESCI 595


RESULT 10
I55476
growth potentiating factor - rat
C;Species: Rattus sp. (rat)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Apr-2001
C;Accession: I55476
R;Nakano, T.; Higashino, K.; Kikuchi, N.; Kishino, J.; Nomura, K.; Fujita, H.; Ohara, O.
J. Biol. Chem. 270, 5702-5705, 1995
A;Title: Vascular smooth muscle cell-derived, Gla-containing growth-potentiating factor
A;Reference number: I55476; MUID:95197586; PMID:7890695
A;Accession: I55476
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-674 <RES>
A;Cross-references: GB:D42148; NID:g1526567; PIDN:BAA07719.1; PID:g893402
C;Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat hom
F;29-89/Domain: Gla domain homology <GLA>
F;117-150/Domain: EGF homology <EG1>
F;157-192/Domain: EGF homology <EG2>
F;198-233/Domain: EGF homology <EG3>
F;239-274/Domain: EGF homology <EG4>
F;308-667/Domain: sex hormone-binding globulin homology <SHB>
F;318-470/Domain: laminin G repeat homology <LGR>

  Query Match             11.8%;  Score 178;  DB 2;  Length 674;
  Best Local Similarity   30.4%;  Pred. No. 1.3e-05;
  Matches   48;  Conservative    8;  Mismatches   66;  Indels   36;  Gaps    6;

Qy          46 QRVYQPFLTTCDGH-------RACSTYRTIYRTAYRRSPGLAPARPRYACC--------- 89
               :| || |     ||         ||      |  :   |      |||  |
Db          44 RRAYQVFEEAKQGHLERECVEEVCSKEEA--REVFENDPETDYFYPRYQECMRKYGRPED 101

Qy          90 --PGWKR-TSGLPGAC--------GAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDV 138
                 | :      ||  |        |  :||    |         | |  || |  |  ||
Db         102 KNPNFATCVKNLPDQCTPNPCDKKGTQLCQDLMGN-------FFCLCKDGWGGRLCDKDV 154

Qy         139 DECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLC 176
               :||| : ||| | | |  ||: | |  | || :|   |
Db         155 NECSQKNGGCSQVCHNKPGSFQCACHSGFSLQSDNKSC 192


RESULT 11
A47221
fibrillin 1 precursor - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 24-Nov-2003
C;Accession: A47221; I54355; S17064; I59574; S17062; S62111; A34198
R;Corson, G.M.; Chalberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.
Genomics 17, 476-484, 1993
A;Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structu
A;Reference number: A47221; MUID:94010947; PMID:7691719
A;Accession: A47221
A;Molecule type: mRNA
A;Residues: 1-337,'T',339-1029 <COR>
A;Cross-references: GB:X63556
R;Pereira, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Pangilinan, T.; Bonad
Hum. Mol. Genet. 2, 961-968, 1993
A;Title: Genomic organization of the sequence coding for fibrillin, the defective gene p
A;Reference number: I54355; MUID:93372860; PMID:8364578
A;Accession: I54355
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 132-3002 <PER>
A;Cross-references: GB:L13923; NID:g306745; PIDN:AAB02036.1; PID:g306746
R;Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.
Nature 352, 334-337, 1991
A;Title: Partial sequence of a candidate gene for the Marfan syndrome.
A;Reference number: S17064; MUID:91304568; PMID:1852207
A;Accession: S17064
A;Molecule type: mRNA
A;Residues: 1030-3002 <MAS>
A;Cross-references: EMBL:X63556
R;Dietz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.
Science 259, 680-683, 1993
A;Title: The skipping of constitutive exons in vivo induced by nonsense mutations.
A;Reference number: I59574; MUID:93157831; PMID:8430317
A;Accession: I59574
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 2217-2288,'I',2290-2325 <RES>
A;Cross-references: GB:S54426; NID:g264860; PIDN:AAB25244.1; PID:g264861
R;Lee, B.; Godfrey, M.; Vitale, E.; Hori, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras,
Nature 352, 330-334, 1991
A;Title: Linkage of Marfan syndrome and a phenotypically related disorder to two diffe
A;Reference number: S17062; MUID:91304567; PMID:1852206
A;Accession: S17062
A;Molecule type: mRNA
A;Residues: 'VLVTVVFIFLSYNKML',944-1444 <LEE1>
A;Cross-references: EMBL:X62008; NID:g31398; PIDN:CAB56534.1; PID:g5924015
A;Accession: S62111
A;Molecule type: protein
A;Residues: 1166-1176,'X',1178-1180,'D',1182-1185 <LEE2>
R;Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.
J. Biol. Chem. 264, 21381-21385, 1989
A;Title: Connective tissue microfibrils. Isolation and characterization of three large
A;Reference number: A34198; MUID:90078246; PMID:2512293
A;Accession: A34198
A;Molecule type: protein
A;Residues: 565-575;1890-1892,'I',1894-1900 <MAD>
C;Comment: Fibrillin is a major component of elastin-associated microfibrils.
C;Genetics:
A;Gene: GDB:FBN1
A;Cross-references: GDB:127115; OMIM:134797; OMIM:154700
A;Map position: 15q21.1-15q21.1
A;Introns: 2236/1; 2258/1; 2297/1
C;Superfamily: fibrillin; EGF homology
C;Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein;
F;1-3002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicte
F;132-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MATC>
F;1332-1367/Domain: EGF homology <EGF>
F;1457-1492/Domain: EGF homology <EGF2>
F;2262-2295/Domain: EGF homology <EGF1>

  Query Match             11.7%;  Score 176.5;  DB 2;  Length 3002;
  Best Local Similarity   22.7%;  Pred. No. 6e-05;
  Matches   59;  Conservative   19;  Mismatches   79;  Indels  103;  Gaps    8;

Qy          26 PGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR---ACSTYRTIYRTAYRRSPGLAPA 82
               ||   | |        : |  |   | : ||   :   :| : |:|     |   | : :
Db         206 PGGNQCIVPICRHSCGDGFCSR---PNMCTCPSGQIAPSCGS-RSIQHCNIRCMNGGSCS 261

Qy          83 RPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQ------- 135
                    |  |:  |      ||  :|:  | ||| || | || |  |: |  |:
Db         262 DDHCLCQKGYIGTH-----CGQPVCESGCLNGGRCVAPNRCACTYGFTGPQCERDYRTGP 316

Qy         136 -----------------------------------------------------------S 136

Db         317 CFTVISNQMCQGQLSGIVCTKQLCCATVGRAWGHPCEMCPAQPHPCRRGFIPNIRTGACQ 376

Qy         137 DVDECSARRGGCP-QRCINTAGSYWCQCWEGHSLSADGTLC-----------VPKGG--- 181
               ||||| |  | |    |||| ||: |:|  || |:     |           : :||
Db         377 DVDECQAIPGLCQGGNCINTVGSFECKCPAGHKLNEVSQKCEDIDECSTIPGICEGGECT 436

Qy         182 ----------PPRVAPNPTG 191
                         ||    :| |
Db         437 NTVSSYFCKCPPGFYTSPDG 456


RESULT 12
```

```
A;Title: The protein encoded by a growth arrest-specific gene (gas6) is a new member of

A;Reference number: A48089; MUID:93330291; PMID:8336730
A;Accession: B48089
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-678 <MAN>
A;Cross-references: GB:L13720; NID:g401766; PIDN:AAA58494.1; PID:g401767
C;Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat hom
F;41-92/Domain: Gla domain homology #status atypical <GLA>
F;120-153/Domain: EGF homology <EG1>
F;160-195/Domain: EGF homology <EG2>
F;201-236/Domain: EGF homology <EG3>
F;242-277/Domain: EGF homology <EG4>
F;311-671/Domain: sex hormone-binding globulin homology <SHB>
F;321-473/Domain: laminin G repeat homology <LGR>

  Query Match             12.6%;  Score 189;  DB 2;  Length 678;
  Best Local Similarity   34.2%;  Pred. No. 2e-06;
  Matches   53;  Conservative    8;  Mismatches   64;  Indels   30;  Gaps    8;

Qy         46 QRVYQPFLTTCDGH-------RACSTYRTIYRTAYRRSPGLAPARPRYACC-----PGWK 93
              :| :| |     ||         ||  |   |  :   |      |||  |       :
Db         47 RRAFQVFEEAKQGHLERECVEELCS--REEAREVFENDPETDYFYPRYLDCINKYGSPYT 104

Qy         94 RTSG-------LPGACGAAICQPPC-RNGGSCVQP--GR--CRCPAGWRGDTCQSDVDEC 141
              : ||       ||  |       || | |    |   |   | | ||| |  |  ||:||
Db        105 KNSGFATCVQNLPDQC----TPNPCDRKGTQACQDLMGNFFCLCKAGWGGRLCDKDVNEC 160

Qy        142 SARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLC 176
              |   ||| | | |  ||: | |  |  ||:||  |
Db        161 SQENGGCLQICHNKPGSFHCSCHSGFELSSDGRTC 195

RESULT 7
A48089
growth arrest-specific protein gas6 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 05-Nov-1999
C;Accession: A48089; S37437
R;Manfioletti, G.; Brancolini, C.; Avanzi, G.; Schneider, C.
Mol. Cell. Biol. 13, 4976-4985, 1993
A;Title: The protein encoded by a growth arrest-specific gene (gas6) is a new member of

A;Reference number: A48089; MUID:93330291; PMID:8336730
A;Accession: A48089
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-673 <MAN>
A;Cross-references: GB:X59846; NID:g407060; PIDN:CAA42507.1; PID:g407061
A;Note: authors translated the codon CCC for residue 424 as Ile
C;Genetics:
A;Gene: gas6
C;Superfamily: plasma protein S; EGF homology; Gla domain homology; laminin G repeat hom
F;38-89/Domain: Gla domain homology #status atypical <GLA>
F;117-150/Domain: EGF homology <EG1>
F;157-192/Domain: EGF homology <EG2>
F;198-233/Domain: EGF homology <EG3>
F;239-274/Domain: EGF homology <EG4>
F;308-666/Domain: sex hormone-binding globulin homology <SHB>
F;318-470/Domain: laminin G repeat homology <LGR>

  Query Match             12.2%;  Score 184;  DB 2;  Length 673;
  Best Local Similarity   31.6%;  Pred. No. 4.6e-06;
  Matches   50;  Conservative    8;  Mismatches   64;  Indels   36;  Gaps    6;

Qy         46 QRVYQPFLTTCDGH-------RACSTYRTIYRTAYRRSPGLAPARPRYACC--------- 89
              :| || |     ||         ||      |  :   |      |||  |
Db         44 RRAYQVFEEAKQGHLERECVEEVCSKEEA--REVFENDPETEYFYPRYQECMRKYGRPEE 101

Qy         90 --PGW-KRTSGLPGAC--------GAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDV 138
                | : |    ||  |        |  |||    |         | |  || |  |  ||
Db        102 KNPDFAKCVQNLPDQCTPNPCDKKGTHICQDLMGN-------FFCVCTDGWGGRLCDKDV 154
```

```
Qy        139 DECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLC 176
              :||  : ||| | | |  ||: | |  | ||::||  |
Db        155 NECVQKNGGCSQVCHNKPGSFQCACHSGFSLASDGQTC 192

RESULT 8
A57278
fibrillin-2 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 24-Nov-2003
C;Accession: A57278
R;Zhang, H.; Hu, W.; Ramirez, F.
J. Cell Biol. 129, 1165-1176, 1995
A;Title: Developmental expression of fibrillin genes suggests heterogeneity of extrace
A;Reference number: A57278; MUID:95263670; PMID:7744963
A;Accession: A57278
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2907 <ZHA>
A;Cross-references: GB:L39790; NID:g762830; PIDN:AAA74908.1; PID:g762831
C;Superfamily: fibrillin; EGF homology
F;1239-1274/Domain: EGF homology <EGF1>
F;2488-2523/Domain: EGF homology <EGF>

  Query Match             12.2%;  Score 183.5;  DB 2;  Length 2907;
  Best Local Similarity   23.9%;  Pred. No. 1.8e-05;
  Matches   60;  Conservative   19;  Mismatches   85;  Indels   87;  Gaps    7;

Qy          2 RGSQEVLLMWLLVLAVGGTEHAY-------RPGRRVCAVRAHGDPVSESFVQRVYQPFLT 54
              || ||:|         |   |:|        ||   | |    :   : |  |   | :
Db         77 RGQQEILRG---PNVCGSRFHSYCCPGWKTLPGGNQCIVPICRNSCGDGFCSR---PNMC 130

Qy         55 TCDGHRACST--YRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCR 112
              ||   :   |   ::| : : |   |   |     |  |:  |      ||  :|:  |:
Db        131 TCSSGQISPTCGRKSIQQCSVRCMNGGTCADDHCQCQKGYIGT-----YCGQPVCENGCQ 185

Qy        113 NGGSCVQPGRCRCPAGWRGDTCQ------------------------------------- 135
              ||| |: | || |  |: |  |:
Db        186 NGGRCIGPNRCACVYGFTGPQCERDYRTGPCFTQVNNQMCQGQLTGIVCTKTLCCATIGR 245

Qy        136 -----------------------------SDVDECSARRGGCP-QRCINTAGSYWCQCWE 165
                                            ||||| |  | |    |||| ||: |:|
Db        246 AWGHPCEMCPAQPQPCRPGFIPNIRTGACQDVDECQAIPGLCQGGNCINTVGSFECRCPA 305

Qy        166 GHSLSADGTLC 176
              ||  |     |
Db        306 GHKQSETTQKC 316

RESULT 9
T10756
Nel-homolog protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: T10756
R;Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Matsuhashi, S.; Kikkawa, U.
submitted to the EMBL Data Library, November 1998
A;Description: Protein kinase C-binding protein.
A;Reference number: Z17122
A;Accession: T10756
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-810 <KUR>
A;Cross-references: EMBL:U48246; NID:g3851179; PID:g3851180
A;Experimental source: strain Sprague-Dawley, brain

  Query Match             12.0%;  Score 180;  DB 2;  Length 810;
  Best Local Similarity   36.5%;  Pred. No. 1.1e-05;
  Matches   35;  Conservative   13;  Mismatches   38;  Indels   10;  Gaps    3;

Qy         88 CCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGG 147
              | ||:       |    | |:  || ||:|| | :| ||:|: |  |: |:|||:
```

```
R;Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A;Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A;Reference number: Z14126; MUID:98360089; PMID:9693030
A;Accession: T13954
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1574 <NAK>
A;Cross-references: EMBL:AB011532; NID:g3449293; PIDN:BAA32462.1; PID:g3449294
A;Experimental source: strain Sprague-Dawley; brain
C;Genetics:
A;Gene: MEGF6

  Query Match             17.0%;  Score 256;  DB 2;  Length 1574;
  Best Local Similarity   37.9%;  Pred. No. 5.2e-11;
  Matches   64;  Conservative   16;  Mismatches   61;  Indels   28;  Gaps    9;

Qy         30 VCAVR-----AHGDPVSESFVQRVYQPFLTTCDGHRAC--STYRTIYRTAYRRSPGLAPA 82
              ||| :      |  |  ::| : |     | |     |     ||:|  :||:      |
Db         43 VCAEQKLTLVGHRQPCVQAFSRIVPVWRRTGCAQQAWCIGQERRTVYYMSYRQVYA-TEA 101

Qy         83 RPRYACCPGWKRTSGLPGA------CGAAI--CQPPCRN--GGSCVQPGRCRCPAGWR-- 130
              |  : ||||| :  |  |       | :|   |: || |  ||       |||| |::
Db        102 RTVFRCCPGWSQKPGQEGCLSDVDECASANGGCEGPCCNTVGGF-----YCRCPPGYQLQ 156

Qy        131 --GDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
                | ||| ||||| |  |||  ||:|| ||| |:|  |  |  ||  |:
Db        157 GDGKTCQ-DVDECRAHNGGCQHRCVNTPGSYLCECKPGFRLHTDGRTCL 204

RESULT 3
T17324
hypothetical protein DKFZp564P2063.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17324
R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A;Reference number: Z18727
A;Accession: T17324
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-558 <DUE>
A;Cross-references: EMBL:AL117610
A;Experimental source: fetal brain; clone DKFZp564P2063
C;Genetics:
A;Note: DKFZp564P2063.1

  Query Match             16.8%;  Score 252.5;  DB 2;  Length 558;
  Best Local Similarity   36.8%;  Pred. No. 3.8e-11;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy         58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
              |: | : :  :  :|  | ||:     : ||| ||:| |   | | | | |:| |:  |  |
Db         26 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 79

Qy        118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
              | | :|||  |: | ||  ||:||  :   |  ||:|| ||| | |  || |  | | ||
Db         80 VGPNKCRCLPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 138

Qy        178 ----------------PKGGPP--------RVAPN 188
                               : ||         |:|||
Db        139 YSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 173

RESULT 4
T27283
hypothetical protein Y64G10A.f - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T27283
R;Ainscough, R.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z20336
A;Accession: T27283
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1620 <WIL>
A;Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f
A;Experimental source: clone Y64G10A
C;Genetics:
A;Gene: CESP:Y64G10A.f
A;Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1;

  Query Match             14.5%;  Score 218;  DB 2;  Length 1620;
  Best Local Similarity   40.4%;  Pred. No. 3.2e-08;
  Matches   42;  Conservative    9;  Mismatches   35;  Indels   18;  Gaps    4;

Qy        102 CGAAICQPPCRNGGSCVQPGR------CRCPAGWRGDTCQSDVDECSARRGGCPQRCINT 155
              | | :    | |||:||          | || |: |  || | :|| |  |||   |:||
Db         79 CSADL----CHNGGTCVPSEHNDNEQVCECPVGFTGAKCQYDANECMANNGGCEHECVNT 134

Qy        156 AGSYWCQCWEGHSLSADGTLC-------VPKGG-PPRVAPNPTG 191
               |:|:|:|| |  || ||  |       |  ||   |   :| |
Db        135 IGTYYCRCWPGFELSGDGNTCSDIDECAVSNGGCSDRCVNSPGG 178

RESULT 5
T09059
notch4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 08-Sep-2002
C;Accession: T09059
R;Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.;
submitted to the EMBL Data Library, October 1997
A;Description: Sequence of the mouse major histocompatibility locus class III region.
A;Reference number: Z16543
A;Accession: T09059
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1964 <ROW>
A;Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564947
C;Genetics:
A;Gene: notch4
A;Map position: 17
A;Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1;
1679/3; 1729/1; 1761/3
C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: receptor; signal transduction
F;514-545/Domain: EGF homology <EGF>

  Query Match             13.0%;  Score 196;  DB 2;  Length 1964;
  Best Local Similarity   34.2%;  Pred. No. 1.6e-06;
  Matches   50;  Conservative   12;  Mismatches   54;  Indels   30;  Gaps    7;

Qy         80 APARPRYACCPGWK-RTSGLPGACGAAICQPPCRNGGSCVQPG---RCRCPAGWRGDTCQ 135
              |  ||: :| |||      |   | |     || ||| |:      :|||| |: | ||:
Db        134 ASGRPQCSCEPGWTGEQCQLRDFCSA----NPCANGGVCLATYPQIQCRCPPGFEGHTCE 189

Qy        136 SDVDECSARRGGCPQ--RCINTAGSYWCQC---WEGHSLSADGTLCVP----KGGPPRVA 186
               |::||    | |||   | || ||| | |    ||         | |     ||  ::
Db        190 RDINECFLEPGPCPQGTSCHNTLGSYQCLCPVGQEGPQCKLRKGACPPGSCLNGGTCQLV 249

Qy        187 PNP-------------TGVDSAMKEE 199
              |               ||:|  |  :
Db        250 PEGHSTFHLCLCPPGFTGLDCEMNPD 275

RESULT 6
B48089
growth arrest-specific protein gas6 - human
C;Species: Homo sapiens (man)
C;Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 05-Nov-1999
C;Accession: B48089
R;Manfioletti, G.; Brancolini, C.; Avanzi, G.; Schneider, C.
Mol. Cell. Biol. 13, 4976-4985, 1993
```

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April 22, 2004, 12:43:15 ; Search time 21 Seconds
                                          (without alignments)
                                          1250.490 Million cell updates/sec

Title:         US-09-978-191A-506
Perfect score: 1505
Sequence:      1 MRGSQEVLLMWLLVLAVGGT..........SEQISFLEEQLGSCSCKKDS 273

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters:      283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      PIR_78:*
             1:  pir1:*
             2:  pir2:*
             3:  pir3:*
             4:  pir4:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                   %
Result          Query
  No.   Score   Match Length DB ID                    Description
------------------------------------------------------------------------------
    1   435.5    28.9    293  2 T09065                hypothetical prote
    2     256    17.0   1574  2 T13954                MEGF6 protein - ra
    3   252.5    16.8    558  2 T17324                hypothetical prote
    4     218    14.5   1620  2 T27283                hypothetical prote
    5     196    13.0   1964  2 T09059                notch4 - mouse
    6     189    12.6    678  2 B48089                growth arrest-spec
    7     184    12.2    673  2 A48089                growth arrest-spec
    8   183.5    12.2   2907  2 A57278                fibrillin-2 precur
    9     180    12.0    810  2 T10756                Nel-homolog protei
   10     178    11.8    674  2 I55476                growth potentiatin
   11   176.5    11.7   3002  2 A47221                fibrillin 1 precur
   12   175.5    11.7   2321  2 S78549                notch3 protein - h
   13     174    11.6    835  2 JP0076                nel protein - chic
   14   173.5    11.5   1247  1 MMHUND                nidogen precursor
   15     173    11.5   2703  1 A24420                notch protein - fr
   16   172.5    11.5   2871  2 A55624                fibrillin-1 precur
   17     172    11.4   2531  2 A46019                notch-1 protein -
   18   171.5    11.4   2871  2 A55567                fibrillin I - bovi
   19     170    11.3    675  1 KXMSS                 plasma protein S p
   20     170    11.3   2918  2 A54105                fibrillin-2 precur
   21   169.5    11.3   2318  2 S45306                notch 3 protein -
   22     169    11.2    833  2 S19087                gene Delta protein
   23     168    11.2    675  1 KXRTS                 plasma protein S p
   24     166    11.0    832  2 A31246                neurogenic protein
   25     166    11.0    880  2 S00670                neurogenic repetit
   26     164    10.9   1429  2 S06434                homeotic protein 1
   27     164    10.9   2524  2 A35844                Xotch protein - Af
   28     164    10.9   3871  2 T22812                hypothetical prote
   29   163.5    10.9   2555  2 A40043                notch protein homo
   30     162    10.8    646  2 S38819                plasma protein S -
   31     161    10.7    387  2 B49175                Motch A protein -
   32     161    10.7    675  1 KXBOS                 plasma protein S p
   33   160.5    10.7   2531  2 S18188                notch protein homo
   34     160    10.6   2437  2 S42612                transmembrane prot
   35   158.5    10.5    407  1 KFBO7                 coagulation factor
   36   158.5    10.5    456  1 KXBO                  protein C (activat
   37     158    10.5   1820  2 A55494                latent transformin
   38     157    10.4    676  1 KXHUS                 plasma protein S p
   39     156    10.4    642  2 S53434                plasma protein S p
   40     156    10.4   1221  2 A49457                fibulin-2 precurso
   41     156    10.4   1408  2 S16148                gene serrate prote
   42   155.5    10.3    466  1 KFHU7                 coagulation factor
   43   155.5    10.3   3635  2 T10053                laminin alpha 5 ch
   44     155    10.3   2471  2 A49128                cell-fate determin
   45     154    10.2    387  2 I38449                extracellular prot

                              ALIGNMENTS

RESULT 1
T09065
hypothetical protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2003
C;Accession: T09065
R;Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.;
submitted to the EMBL Data Library, October 1997
A;Description: Sequence of the mouse major histocompatibility locus class III region.
A;Reference number: Z16543
A;Accession: T09065
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-293 <ROW>
A;Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564953
C;Genetics:
A;Map position: 17
A;Introns: 34/2; 75/2; 112/1; 144/1; 201/1; 228/3; 280/1
F;114-141/Domain: EGF homology <EGF1>
F;148-183/Domain: EGF homology <EGF>

  Query Match             28.9%;  Score 435.5;  DB 2;  Length 293;
  Best Local Similarity   37.3%;  Pred. No. 9e-25;
  Matches  112;  Conservative   36;  Mismatches  101;  Indels   51;  Gaps   14;

Qy          1 MRGSQEVLLMWLLVLAVG-GTE-HAYRPGRRVCAVRAHGDPV--SESFVQRVYQPFLTTC 56
              :||     | : |||  | ||   :::    ||: :    |:  :||: | ||:|:|| |
Db         11 LRG-----LSFFLVLMTGEGTRGGSFKESLGVCSKQTLLVPLRYNESYSQPVYKPYLTLC 65

Qy         57 DGHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGA-AICQPPCRNGG 115
               | | |||||| || |:|      | :    || |||:    |||    |||  || |||
Db         66 AGRRICSTYRTTYRVAWREVRREVP-QTHVVCCQGWKKPH--PGALTCDAICSKPCLNGG 122

Qy        116 SCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTL 175
               |  | || |  || |  |  ||||| |    |   |:|| ||: | |     |  ||
Db        123 VCTGPDRCECAPGWGGKHCHVDVDECRASLTLCSHGCLNTLGSFLCSCPHPLVLGLDGRT 182

Qy        176 CVPKGGPPRVAPNPTGV--------------DSAMKEEVQRLQSRVDLLEEKLQLVLAPL 221
              |   ||||    :||                :  |:: ||   : |:: ||:
Db        183 CA--GGPPE---SPTSASILSVAVREADSEEERALRWEVAELRGRLEKLEQ--------- 228

Qy        222 HSLASQA---LEHGLP-DPGSLLVHSFQQL----GRIDSLSEQISFLEEQLGSCSCKKDS 273
                 |:||   :   ||  |  |      :|     ||:|||:|:  |||:||:|:|: :|
Db        229 --WATQAGAWVRAVLPMPPEELRPEQVAELWGRGDRIESLSDQVLLLEERLGACACEDNS 286

RESULT 2
T13954
MEGF6 protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T13954
```

```
  Query Match             14.4%;  Score 216.5;  DB 4;  Length 537;
  Best Local Similarity   39.3%;  Pred. No. 1.7e-09;
  Matches   48;  Conservative   10;  Mismatches   35;  Indels   29;  Gaps    6;

Qy         91 GWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQ 150
              ||:| |   | |  | |:| |:  | || | :|||  |: | ||  ||:||  :   |
Db          1 GWRRNS--KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQH 56

Qy        151 RCINTAGSYWCQCWEGHSLSADGTLCV----------------PKGGPP--------RVA 186
              ||:|| ||| | |  || |  | | ||                 : ||         |:|
Db         57 RCVNTHGSYKCFCLSGHMLMPDAT-CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLA 115

Qy        187 PN 188
              ||
Db        116 PN 117


RESULT 13
US-09-249-697A-3
; Sequence 3, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
;  TITLE OF INVENTION:  LIVER SPLEEN
;  FILE REFERENCE: 24011-727
;  CURRENT APPLICATION NUMBER: US/09/249,697A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 100
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-249-697A-3

  Query Match             14.3%;  Score 215.5;  DB 4;  Length 100;
  Best Local Similarity   48.3%;  Pred. No. 3.1e-10;
  Matches   42;  Conservative    8;  Mismatches   32;  Indels    5;  Gaps    4;

Qy         91 GWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQ 150
              ||:| |   | |  | |:| |:  | || | :|||  |: | ||  ||:||  :   |
Db          1 GWRRNS--KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQH 56

Qy        151 RCINTAGSYWCQCWEGHSLSADGTLCV 177
              ||:|| ||| | |  || |  | | ||
Db         57 RCVNTHGSYKCFCLSGHMLMPDAT-CV 82


RESULT 14
US-09-363-316B-3
; Sequence 3, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 100
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-363-316B-3

  Query Match             14.3%;  Score 215.5;  DB 4;  Length 100;
  Best Local Similarity   48.3%;  Pred. No. 3.1e-10;
  Matches   42;  Conservative    8;  Mismatches   32;  Indels    5;  Gaps    4;

Qy         91 GWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQ 150
              ||:| |   | |  | |:| |:  | || | :|||  |: | ||  ||:||  :   |
Db          1 GWRRNS--KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQH 56

Qy        151 RCINTAGSYWCQCWEGHSLSADGTLCV 177
              ||:|| ||| | |  || |  | | ||
Db         57 RCVNTHGSYKCFCLSGHMLMPDAT-CV 82


RESULT 15
US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
;  APPLICANT: Kitajewski, Jan
;  APPLICANT:  Uyttendaele, Hendrik
;  TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
;  FILE REFERENCE: 53863-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/467,997
;  CURRENT FILING DATE:  1999-12-20
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 1964
;   TYPE: PRT
;   ORGANISM: mouse
US-09-467-997-1

  Query Match             13.0%;  Score 196;  DB 4;  Length 1964;
  Best Local Similarity   34.2%;  Pred. No. 3.2e-07;
  Matches   50;  Conservative   12;  Mismatches   54;  Indels   30;  Gaps    7;

Qy         80 APARPRYACCPGWK-RTSGLPGACGAAICQPPCRNGGSCVQPG---RCRCPAGWRGDTCQ 135
              |  ||: :| |||      |   | |     || ||| |:      :|||| |: | ||:
Db        134 ASGRPQCSCEPGWTGEQCQLRDFCSA----NPCANGGVCLATYPQIQCRCPPGFEGHTCE 189

Qy        136 SDVDECSARRGGCPQ--RCINTAGSYWCQC---WEGHSLSADGTLCVP----KGGPPRVA 186
               |::||    | |||   | || ||| | |    ||         | |     ||  ::
Db        190 RDINECFLEPGPCPQGTSCHNTLGSYQCLCPVGQEGPQCKLRKGACPPGSCLNGGTCQLV 249

Qy        187 PNP-------------TGVDSAMKEE 199
              |               ||:|  |  :
Db        250 PEGHSTFHLCLCPPGFTGLDCEMNPD 275


Search completed: April 22, 2004, 12:47:57
Job time : 23 secs
```

```
;  PRIOR FILING DATE: 1999-11-30
;  PRIOR APPLICATION NUMBER: PCT/US99/28564
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/28565
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/30095
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: PCT/US99/30911
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US99/30999
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US00/00219
;  PRIOR FILING DATE: 2000-01-05
;  NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 315
;   LENGTH: 509
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-902-775A-315

  Query Match             14.8%;  Score 223;  DB 4;  Length 509;
  Best Local Similarity   42.4%;  Pred. No. 5e-10;
  Matches   42;  Conservative   13;  Mismatches   40;  Indels    4;  Gaps    3;

Qy         76 SPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQ 135
              | ||     |  || || | |   | |   :||| |:: | |: | :|:|  |: | || 
Db         32 SIGLCRYGGRIDCCWGWARQSW--GQC-QPVCQPRCKH-GECIGPNKCKCHPGYAGKTCN 87

Qy        136 SDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGT 174
               |::||  :   |  ||:|| ||| | |  |: |  ||:
Db         88 QDLNECGLKPRPCKHRCMNTYGSYKCYCLNGYMLMPDGS 126

RESULT 10
US-09-363-316B-18
; Sequence 18, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 18
;   LENGTH: 502
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (501-502)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-18

  Query Match             14.4%;  Score 216.5;  DB 4;  Length 502;
  Best Local Similarity   39.3%;  Pred. No. 1.6e-09;
  Matches   48;  Conservative   10;  Mismatches   35;  Indels   29;  Gaps    6;

Qy         91 GWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQ 150
              ||:| |   | |  |:| | |:  | || | :|||  |: | ||  ||:||  :   |  
Db          1 GWRRNS--KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQH 56

Qy        151 RCINTAGSYWCQCWEGHSLSADGTLCV----------------PKGGPP--------RVA 186
              ||:|| ||| | |  || |  | | ||                 : ||         |:|
Db         57 RCVNTHGSYKCFCLSGHMLMPDAT-CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLA 115

Qy        187 PN 188
```

```
              ||
Db        116 PN 117

RESULT 11
US-09-249-697A-4
; Sequence 4, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
;  TITLE OF INVENTION:  LIVER SPLEEN
;  FILE REFERENCE: 24011-727
;  CURRENT APPLICATION NUMBER: US/09/249,697A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 537
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(537)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-249-697A-4

  Query Match             14.4%;  Score 216.5;  DB 4;  Length 537;
  Best Local Similarity   39.3%;  Pred. No. 1.7e-09;
  Matches   48;  Conservative   10;  Mismatches   35;  Indels   29;  Gaps    6;

Qy         91 GWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQ 150
              ||:| |   | |  |:| | |:  | || | :|||  |: | ||  ||:||  :   |  
Db          1 GWRRNS--KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQH 56

Qy        151 RCINTAGSYWCQCWEGHSLSADGTLCV----------------PKGGPP--------RVA 186
              ||:|| ||| | |  || |  | | ||                 : ||         |:|
Db         57 RCVNTHGSYKCFCLSGHMLMPDAT-CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLA 115

Qy        187 PN 188
              ||
Db        116 PN 117

RESULT 12
US-09-363-316B-4
; Sequence 4, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
;   LENGTH: 537
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (503)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-4
```

```
RESULT 8
US-09-905-125A-315
; Sequence 315, Application US/09905125A
; Patent No. 6664376
; GENERAL INFORMATION:
;  APPLICANT: Genentech, Inc.
;  APPLICANT:  Ashkenazi, Avi
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, A.
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth, J.
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Mather, Jennie P.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William, I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: 10466-14
;  CURRENT APPLICATION NUMBER: US/09/905,125A
;  CURRENT FILING DATE:  2001-07-12
;  PRIOR APPLICATION NUMBER: PCT/US00/04414
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: US 60/143,048
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: US 60/145,698
;  PRIOR FILING DATE: 1999-07-26
;  PRIOR APPLICATION NUMBER: US 60/146,222
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: PCT/US99/20594
;  PRIOR FILING DATE: 1999-09-08
;  PRIOR APPLICATION NUMBER: PCT/US99/20944
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: PCT/US99/21090
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/21547
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/23089
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: PCT/US99/28214
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: PCT/US99/28313
;  PRIOR FILING DATE: 1999-11-30
;  PRIOR APPLICATION NUMBER: PCT/US99/28564
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/28565
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/30095
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: PCT/US99/30911
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US99/30999
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US00/00219
;  PRIOR FILING DATE: 2000-01-05
;  NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 315
;   LENGTH: 509
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-905-125A-315

  Query Match             14.8%;  Score 223;  DB 4;  Length 509;
  Best Local Similarity   42.4%;  Pred. No. 5e-10;
  Matches   42;  Conservative   13;  Mismatches   40;  Indels    4;  Gaps    3;

Qy         76 SPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQ 135
              | ||     |  || || | |   | |   :||| |:: | |: | :|:|  |: | ||
Db         32 SIGLCRYGGRIDCCWGWARQSW--GQC-QPVCQPRCKH-GECIGPNKCKCHPGYAGKTCN 87

Qy        136 SDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGT 174
               |::||  :   |  ||:|| ||| | |  |: |  ||:
Db         88 QDLNECGLKPRPCKHRCMNTYGSYKCYCLNGYMLMPDGS 126


RESULT 9
US-09-902-775A-315
; Sequence 315, Application US/09902775A
; Patent No. 6686451
; GENERAL INFORMATION:
;  APPLICANT: Genentech, Inc.
;  APPLICANT:  Ashkenazi, Avi
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, A.
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth, J.
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Mather, Jennie P.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William, I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: 10466-14
;  CURRENT APPLICATION NUMBER: US/09/902,775A
;  CURRENT FILING DATE:  2001-07-10
;  PRIOR APPLICATION NUMBER: PCT/US00/04414
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: US 60/143,048
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: US 60/145,698
;  PRIOR FILING DATE: 1999-07-26
;  PRIOR APPLICATION NUMBER: US 60/146,222
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: PCT/US99/20594
;  PRIOR FILING DATE: 1999-09-08
;  PRIOR APPLICATION NUMBER: PCT/US99/20944
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: PCT/US99/21090
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/21547
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/23089
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: PCT/US99/28214
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: PCT/US99/28313
```

```
  Query Match             16.7%;  Score 251.5;  DB 4;  Length 553;
  Best Local Similarity   36.8%;  Pred. No. 2.9e-12;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy          58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
               |: | : :  :  :|  | ||:     : ||| ||:| |   | |  | |:| |:  | |
Db          21 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 74

Qy         118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
               | | :|||  |: | ||  ||:||  :   |  ||:|| ||| | |  || |  | | ||
Db          75 VGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 133

Qy         178 ----------------PKGGPP--------RVAPN 188
                                : ||         |:|||
Db         134 NSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168

RESULT 6
US-09-363-316B-24
; Sequence 24, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-363-316B-24

  Query Match             16.7%;  Score 251.5;  DB 4;  Length 553;
  Best Local Similarity   36.8%;  Pred. No. 2.9e-12;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy          58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
               |: | : :  :  :|  | ||:     : ||| ||:| |   | |  | |:| |:  | |
Db          21 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 74

Qy         118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
               | | :|||  |: | ||  ||:||  :   |  ||:|| ||| | |  || |  | | ||
Db          75 VGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 133

Qy         178 ----------------PKGGPP--------RVAPN 188
                                : ||         |:|||
Db         134 NSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168

RESULT 7
US-09-907-794A-315
; Sequence 315, Application US/09907794A
; Patent No. 6635468
; GENERAL INFORMATION:
;  APPLICANT: Genentech, Inc.
;  APPLICANT:  Ashkenazi, Avi
;  APPLICANT:  Botstein, David
;  APPLICANT:  Desnoyers, Luc
;  APPLICANT:  Eaton, Dan L.
;  APPLICANT:  Ferrara, Napoleone
;  APPLICANT:  Filvaroff, Ellen
;  APPLICANT:  Fong, Sherman
;  APPLICANT:  Gao, Wei-Qiang
;  APPLICANT:  Gerber, Hanspeter
;  APPLICANT:  Gerritsen, Mary E.
;  APPLICANT:  Goddard, A.
;  APPLICANT:  Godowski, Paul J.
;  APPLICANT:  Grimaldi, Christopher J.
;  APPLICANT:  Gurney, Austin L.
;  APPLICANT:  Hillan, Kenneth, J.
;  APPLICANT:  Kljavin, Ivar J.
;  APPLICANT:  Mather, Jennie P.
;  APPLICANT:  Pan, James
;  APPLICANT:  Paoni, Nicholas F.
;  APPLICANT:  Roy, Margaret Ann
;  APPLICANT:  Stewart, Timothy A.
;  APPLICANT:  Tumas, Daniel
;  APPLICANT:  Williams, P. Mickey
;  APPLICANT:  Wood, William, I.
;  TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
;  TITLE OF INVENTION:  Acids Encoding the Same
;  FILE REFERENCE: 10466-14
;  CURRENT APPLICATION NUMBER: US/09/907,794A
;  CURRENT FILING DATE:  2001-07-17
;  PRIOR APPLICATION NUMBER: PCT/US00/04414
;  PRIOR FILING DATE: 2000-02-22
;  PRIOR APPLICATION NUMBER: US 60/143,048
;  PRIOR FILING DATE: 1999-07-07
;  PRIOR APPLICATION NUMBER: US 60/145,698
;  PRIOR FILING DATE: 1999-07-26
;  PRIOR APPLICATION NUMBER: US 60/146,222
;  PRIOR FILING DATE: 1999-07-28
;  PRIOR APPLICATION NUMBER: PCT/US99/20594
;  PRIOR FILING DATE: 1999-09-08
;  PRIOR APPLICATION NUMBER: PCT/US99/20944
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: PCT/US99/21090
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/21547
;  PRIOR FILING DATE: 1999-09-15
;  PRIOR APPLICATION NUMBER: PCT/US99/23089
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: PCT/US99/28214
;  PRIOR FILING DATE: 1999-11-29
;  PRIOR APPLICATION NUMBER: PCT/US99/28313
;  PRIOR FILING DATE: 1999-11-30
;  PRIOR APPLICATION NUMBER: PCT/US99/28564
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/28565
;  PRIOR FILING DATE: 1999-12-02
;  PRIOR APPLICATION NUMBER: PCT/US99/30095
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: PCT/US99/30911
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US99/30999
;  PRIOR FILING DATE: 1999-12-20
;  PRIOR APPLICATION NUMBER: PCT/US00/00219
;  PRIOR FILING DATE: 2000-01-05
;  NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 315
;   LENGTH: 509
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-907-794A-315

  Query Match             14.8%;  Score 223;  DB 4;  Length 509;
  Best Local Similarity   42.4%;  Pred. No. 5e-10;
  Matches   42;  Conservative   13;  Mismatches   40;  Indels    4;  Gaps    3;

Qy          76 SPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQ 135
               | ||     |  || || | |   | |   :||| |:: | |: | :|:|  |: | ||
Db          32 SIGLCRYGGRIDCCWGWARQSW--GQC-QPVCQPRCKH-GECIGPNKCKCHPGYAGKTCN 87

Qy         136 SDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGT 174
                |::||  :   |  ||:|| ||| | |  |: |  ||:
Db          88 QDLNECGLKPRPCKHRCMNTYGSYKCYCLNGYMLMPDGS 126
```

```
; Patent No. 6392018
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
;  TITLE OF INVENTION:  LIVER SPLEEN
;  FILE REFERENCE: 24011-727
;  CURRENT APPLICATION NUMBER: US/09/249,697A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(553)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-249-697A-6

  Query Match             16.9%;  Score 254.5;  DB 4;  Length 553;
  Best Local Similarity   34.5%;  Pred. No. 1.7e-12;
  Matches   60;  Conservative   21;  Mismatches   62;  Indels   31;  Gaps    7;

Qy         39 PVSESFVQRVYQPFLTTCDGHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGL 98
              |:  |    :  |::    |: | : :  :  :|  | ||:     : ||| ||:| |
Db          2 PLPWSLALPLLLPWVAGGFGNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS-- 57

Qy         99 PGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGS 158
               | |  | |:| |:  | || | :|||  |: | ||  ||:||  :   |  ||:|| ||
Db         58 KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGS 115

Qy        159 YWCQCWEGHSLSADGTLCV----------------PKGGPP--------RVAPN 188
              | | |  || |  | | ||                 : ||         |:|||
Db        116 YKCFCLSGHMLMPDAT-CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168

RESULT 3
US-09-363-316B-6
; Sequence 6, Application US/09363316B
; Patent No. 6392019
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: EGF MOTIF PROTEIN MATERIALS AND METHODS
;  FILE REFERENCE: 28110/35852
;  CURRENT APPLICATION NUMBER: US/09/363,316B
;  CURRENT FILING DATE:  1999-07-28
;  PRIOR APPLICATION NUMBER: US 09/249,697
;  PRIOR FILING DATE: 1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (357)
;   OTHER INFORMATION: Xaa = any amino acid
US-09-363-316B-6

  Query Match             16.9%;  Score 254.5;  DB 4;  Length 553;
  Best Local Similarity   34.5%;  Pred. No. 1.7e-12;
  Matches   60;  Conservative   21;  Mismatches   62;  Indels   31;  Gaps    7;

Qy         39 PVSESFVQRVYQPFLTTCDGHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGL 98
              |:  |    :  |::    |: | : :  :  :|  | ||:     : ||| ||:| |
Db          2 PLPWSLALPLLLPWVAGGFGNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS-- 57

Qy         99 PGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGS 158
               | |  | |:| |:  | || | :|||  |: | ||  ||:||  :   |  ||:|| ||
Db         58 KGVC-EATCEPGCK-FGECVGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGS 115

Qy        159 YWCQCWEGHSLSADGTLCV----------------PKGGPP--------RVAPN 188
              | | |  || |  | | ||                 : ||         |:|||
Db        116 YKCFCLSGHMLMPDAT-CVNSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168

RESULT 4
US-09-312-283C-389
; Sequence 389, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 389
;   LENGTH: 284
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-312-283C-389

  Query Match             16.8%;  Score 252.5;  DB 4;  Length 284;
  Best Local Similarity   36.3%;  Pred. No. 1.1e-12;
  Matches   53;  Conservative   15;  Mismatches   51;  Indels   27;  Gaps    4;

Qy         66 RTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRC 125
              |  :  |     ||:     : ||| |||| |   | |  |:|:| |:  | || | :|||
Db         47 RDYWLPALAHQPGVCHYGTKTACCYGWKRNS--KGVC-EAVCEPRCK-FGECVGPNKCRC 102

Qy        126 PAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSL---------------- 169
                |: | ||  ||:||: :   |  ||:|| ||| | |  || |
Db        103 FPGYTGKTCSQDVNECAFKPRPCQHRCVNTHGSYKCFCLSGHMLLPDATCSNSRTCARIN 162

Qy        170 -------SADGTLCVPKGGPPRVAPN 188
                     :|:|  ||      |: ||
Db        163 CQYSCEDTAEGPRCVCPSSGLRLGPN 188

RESULT 5
US-09-249-697A-19
; Sequence 19, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
;  TITLE OF INVENTION:  LIVER SPLEEN
;  FILE REFERENCE: 24011-727
;  CURRENT APPLICATION NUMBER: US/09/249,697A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-249-697A-19
```

```
                           GenCore version 5.1.6
                  Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          April 22, 2004, 12:44:11 ; Search time 22 Seconds
                                            (without alignments)
                                            640.631 Million cell updates/sec

Title:           US-09-978-191A-506
Perfect score:   1505
Sequence:        1 MRGSQEVLLMWLLVLAVGGT..........SEQISFLEEQLGSCSCKKDS 273

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:      389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
              1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
              2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
              3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
              4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
              5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
              6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                      %
Result              Query
   No.    Score     Match Length DB ID                         Description
------------------------------------------------------------------------------
     1     1180     78.4    278  4  US-09-724-864-52            Sequence 52, Appl
     2    254.5     16.9    553  4  US-09-249-697A-6            Sequence 6, Appli
     3    254.5     16.9    553  4  US-09-363-316B-6            Sequence 6, Appli
     4    252.5     16.8    284  4  US-09-312-283C-389          Sequence 389, App
     5    251.5     16.7    553  4  US-09-249-697A-19           Sequence 19, Appl
     6    251.5     16.7    553  4  US-09-363-316B-24           Sequence 24, Appl
     7      223     14.8    509  4  US-09-907-794A-315          Sequence 315, App
     8      223     14.8    509  4  US-09-905-125A-315          Sequence 315, App
     9      223     14.8    509  4  US-09-902-775A-315          Sequence 315, App
    10    216.5     14.4    502  4  US-09-363-316B-18           Sequence 18, Appl
    11    216.5     14.4    537  4  US-09-249-697A-4            Sequence 4, Appli
    12    216.5     14.4    537  4  US-09-363-316B-4            Sequence 4, Appli
    13    215.5     14.3    100  4  US-09-249-697A-3            Sequence 3, Appli
    14    215.5     14.3    100  4  US-09-363-316B-3            Sequence 3, Appli
    15      196     13.0   1964  4  US-09-467-997-1             Sequence 1, Appli
    16      189     12.6    678  1  US-08-282-141-2             Sequence 2, Appli
    17      189     12.6    678  1  US-08-435-434-2             Sequence 2, Appli
    18      189     12.6    678  1  US-08-435-436-2             Sequence 2, Appli
    19      189     12.6    678  2  US-08-438-863-2             Sequence 2, Appli
    20      189     12.6    678  2  US-08-438-864-2             Sequence 2, Appli
    21      189     12.6    678  3  US-08-438-862-2             Sequence 2, Appli
    22      189     12.6    678  3  US-08-628-747-2             Sequence 2, Appli
    23      189     12.6    678  3  US-08-402-253-2             Sequence 2, Appli
    24      189     12.6    678  3  US-08-443-866B-2            Sequence 2, Appli
    25      184     12.2    673  1  US-08-282-141-3             Sequence 3, Appli
    26      184     12.2    673  1  US-08-435-434-1             Sequence 1, Appli
    27      184     12.2    673  1  US-08-435-436-1             Sequence 1, Appli
```

```
    28      184     12.2    673  2  US-08-438-863-1             Sequence 1, Appli
    29      184     12.2    673  2  US-08-438-864-1             Sequence 1, Appli
    30      184     12.2    673  3  US-08-438-862-1             Sequence 1, Appli
    31      184     12.2    673  3  US-08-628-747-1             Sequence 1, Appli
    32      184     12.2    673  3  US-08-402-253-1             Sequence 1, Appli
    33      184     12.2    673  3  US-08-443-866B-1            Sequence 1, Appli
    34    182.5     12.1    816  2  US-08-820-170A-37           Sequence 37, Appl
    35    182.5     12.1    816  3  US-09-055-699-37            Sequence 37, Appl
    36    182.5     12.1    816  3  US-09-273-565-37            Sequence 37, Appl
    37    182.5     12.1    816  4  US-09-565-538-37            Sequence 37, Appl
    38    182.5     12.1    816  4  US-09-661-468-37            Sequence 37, Appl
    39    182.5     12.1    816  4  US-09-976-165-37            Sequence 37, Appl
    40      182     12.1    810  2  US-08-820-170A-34           Sequence 34, Appl
    41      182     12.1    810  3  US-09-055-699-34            Sequence 34, Appl
    42      182     12.1    810  3  US-09-273-565-34            Sequence 34, Appl
    43      182     12.1    810  4  US-09-565-538-34            Sequence 34, Appl
    44      182     12.1    810  4  US-09-661-468-34            Sequence 34, Appl
    45      182     12.1    810  4  US-09-976-165-34            Sequence 34, Appl

                                ALIGNMENTS


RESULT 1
US-09-724-864-52
; Sequence 52, Application US/09724864
; Patent No. 6380362
; GENERAL INFORMATION:
;  APPLICANT:  Watson, James D
;  APPLICANT:  Murison, James G.
;  TITLE OF INVENTION:  Polynucleotides, polypeptides expressed
;  TITLE OF INVENTION:  by the polynucleotides and methods for their use.
;  FILE REFERENCE:  11000.1050U1
;  CURRENT APPLICATION NUMBER: US/09/724,864
;  CURRENT FILING DATE:  2000-11-28
;  PRIOR APPLICATION NUMBER:  U.S. No. 6380362 60/171,678
;  PRIOR FILING DATE:  1999-12-23
;  NUMBER OF SEQ ID NOS:  72
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 52
;   LENGTH: 278
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-724-864-52

  Query Match             78.4%;  Score 1180;  DB 4;  Length 278;
  Best Local Similarity   77.7%;  Pred. No. 1.4e-86;
  Matches  213;  Conservative   19;  Mismatches   40;  Indels    2;  Gaps    2;

Qy          1 MRGSQEVLLMWLLVLAVGG-TEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGH 59
              | || |:|: | ||||  | ||| ||| |||| |   |  :||:||||||||:|||||||
Db          4 MWGSGELLVAWFLVLAADGTTEHVYRPSRRVCTVGISGGSISETFVQRVYQPYLTTCDGH 63

Qy         60 RACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQ 119
              |||||||||||||||||||: ||||||||||||||||||||||||||||||| |||||::
Db         64 RACSTYRTIYRTAYRRSPGVTPARPRYACCPGWKRTSGLPGACGAAICQPPCGNGGSCIR 123

Qy        120 PGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPK 179
              || |||| ||:|||||:||||||     |||||:|| |||||| ||| | ||||| |: |
Db        124 PGHCRCPVGWQGDTCQTDVDECSTGEASCPQRCVNTVGSYWCQGWEGQSPSADGTRCLSK 183

Qy        180 GGPPRVAPNPT-GVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGS 238
               ||  |||||| ||||  :||| |||:|||:||:||||||||||||||:: |||| ||||
Db        184 EGPSPVAPNPTAGVDSMAREEVYRLQARVDVLEQKLQLVLAPLHSLASRSTEHGLQDPGS 243

Qy        239 LLVHSFQQLGRIDSLSEQISFLEEQLGSCSCKKD 272
              || |||||| ||||||||:||||| |||||||||
Db        244 LLAHSFQQLDRIDSLSEQVSFLEEHLGSCSCKKD 277


RESULT 2
US-09-249-697A-6
; Sequence 6, Application US/09249697A
```

```
KW   colon tumour; breast tumour; prostate tumour; rectal tumour;
KW   liver tumour; bone disorder; cartilage disorder; sports injury;
KW   arthritis; wound.
XX
OS   Homo sapiens.
XX
PN   US2003045687-A1.
XX
PD   06-MAR-2003.
XX
PF   12-AUG-2002; 2002US-00218631.
XX
PR   01-JUN-2001; 2001WO-US017800.
PR   29-JUN-2001; 2001WO-US021066.
PR   09-APR-2002; 2002US-00119480.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Desnoyers L,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Smith V,  Stephan JF,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-512315/48.
DR   N-PSDB; ACD68681.
XX
PT   New genes, and its encoded secreted and transmembrane polypeptides,
PT   useful for stimulating Tumor Necrosis Factor alpha, or chondrocyte or
PT   pericyte proliferation, especially for treating lung tumors, arthritis or
PT   wounds in a mammal.
XX
PS   Claim 11; Fig 178; 314pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule comprising a
CC   sequence with at least 80% identity to: (a) a nucleotide encoding any of
CC   122 PRO (secreted and transmembrane) polypeptides whose sequences are
CC   fully defined in the specification; or (b) any of 122 nucleotide
CC   sequences having e.g. 4834, 2504 or 1759 bp fully defined in the
CC   specification; or the full length coding sequence of any these 122
CC   nucleotide sequences. The PRO polypeptides or polynucleotides are useful
CC   as pharmaceuticals, diagnostics, biosensors or bioreactors. These are
CC   particularly useful for detecting tumours (e.g. lung tumour, colon
CC   tumour, breast tumour, prostate tumour, rectal tumour, or liver tumour)
CC   in a mammal, for stimulating the release of TNF-alpha from human blood,
CC   for stimulating the proliferation or differentiation of chondrocyte
CC   cells, for stimulating proliferation of pericyte cells, or for modulating
CC   normal human dermal fibroblast proliferation. The PRO nucleic acid or
CC   polypeptide is also useful for treating tumours or various bone and/or
CC   cartilage disorders (e.g. sports injuries or arthritis), or wounds. The
CC   PRO polypeptides are useful in drug screening, particularly as targets
CC   for therapeutic intervention in these diseases, and in the diagnostic
CC   determination of the presence of these diseases. The PRO polypeptides are
CC   also useful as molecular weight markers, or for chromosome
CC   identification. The PRO genes are useful as hybridisation probes, or for
CC   screening libraries of human cDNA, genomic DNA or mRNA. The PRO genes may
CC   also be used in gene therapy, particularly for replacing a defective
CC   gene. This is the amino acid sequence of a novel human secreted and
CC   transmembrane PRO polypeptide
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 6;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


Search completed: April 22, 2004, 12:45:20
Job time : 63 secs
```

```
PR   13-MAY-1998;   98US-0085339P.
PR   15-MAY-1998;   98US-0085573P.
PR   15-MAY-1998;   98US-0085579P.
PR   15-MAY-1998;   98US-0085580P.
PR   15-MAY-1998;   98US-0085582P.
PR   15-MAY-1998;   98US-0085689P.
PR   15-MAY-1998;   98US-0085697P.
PR   15-MAY-1998;   98US-0085700P.
PR   15-MAY-1998;   98US-0085704P.
PR   18-MAY-1998;   98US-0086023P.
PR   22-MAY-1998;   98US-0086392P.
PR   22-MAY-1998;   98US-0086414P.
PR   22-MAY-1998;   98US-0086430P.
PR   22-MAY-1998;   98US-0086486P.
PR   28-MAY-1998;   98US-0087098P.
PR   28-MAY-1998;   98US-0087106P.
PR   28-MAY-1998;   98US-0087208P.
PR   26-JUN-1998;   98US-00105413.
PR   26-JUN-1998;   98US-0090863P.
PR   26-JUN-1998;   98US-0091010P.
PR   01-JUL-1998;   98US-0091359P.
PR   30-JUL-1998;   98US-0094651P.
PR   11-SEP-1998;   98US-0100038P.
PR   07-OCT-1998;   98US-00168978.
PR   07-OCT-1998;   98WO-US021141.
PR   02-NOV-1998;   98US-00184216.
PR   06-NOV-1998;   98US-00187368.
PR   20-NOV-1998;   98US-0109304P.
PR   20-NOV-1998;   98WO-US024855.
PR   07-DEC-1998;   98US-00202054.
PR   22-DEC-1998;   98US-00218517.
PR   22-DEC-1998;   98US-0113296P.
PR   23-DEC-1998;   98US-0113621P.
PR   05-JAN-1999;   99WO-US000106.
PR   05-MAR-1999;   99US-00254465.
PR   08-MAR-1999;   99WO-US005028.
PR   10-MAR-1999;   99US-00265686.
PR   10-MAR-1999;   99WO-US005190.
PR   12-MAR-1999;   99US-00267213.
PR   12-MAR-1999;   99US-0123957P.
PR   29-MAR-1999;   99US-0126773P.
PR   12-APR-1999;   99US-00284291.
PR   21-APR-1999;   99US-0130232P.
PR   26-APR-1999;   99US-0131022P.
PR   28-APR-1999;   99US-0131445P.
PR   14-MAY-1999;   99US-00311832.
PR   14-MAY-1999;   99US-0134287P.
PR   14-MAY-1999;   99WO-US010733.
PR   02-JUN-1999;   99WO-US012252.
PR   16-JUN-1999;   99US-0139557P.
PR   23-JUN-1999;   99US-0141037P.
PR   07-JUL-1999;   99US-0142680P.
PR   26-JUL-1999;   99US-0145698P.
PR   28-JUL-1999;   99US-0146222P.
PR   25-AUG-1999;   99US-00380137.
PR   25-AUG-1999;   99US-00380138.
PR   25-AUG-1999;   99US-00380142.
PR   29-OCT-1999;   99US-0162506P.
PR   30-NOV-1999;   99WO-US028313.
PR   02-DEC-1999;   99WO-US028551.
PR   02-DEC-1999;   99WO-US028565.
PR   16-DEC-1999;   99WO-US030095.
PR   30-DEC-1999;   99WO-US031243.
PR   30-DEC-1999;   99WO-US031274.
PR   05-JAN-2000; 2000WO-US000219.
PR   06-JAN-2000; 2000WO-US000277.
PR   06-JAN-2000; 2000WO-US000376.
PR   11-FEB-2000; 2000WO-US003565.
PR   18-FEB-2000; 2000WO-US004341.
PR   24-FEB-2000; 2000WO-US005004.
PR   02-MAR-2000; 2000WO-US005841.
PR   10-MAR-2000; 2000WO-US006319.
PR   21-MAR-2000; 2000WO-US007532.
```

```
PR   30-MAR-2000; 2000WO-US008439.
PR   17-MAY-2000; 2000WO-US013705.
PR   22-MAY-2000; 2000WO-US014042.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   28-JUL-2000; 2000WO-US020710.
PR   24-AUG-2000; 2000WO-US023328.
PR   08-NOV-2000; 2000US-00709238.
PR   27-NOV-2000; 2000US-00723749.
PR   01-DEC-2000; 2000WO-US032678.
PR   20-DEC-2000; 2000US-00747259.
PR   20-DEC-2000; 2000WO-US034956.
PR   28-FEB-2001; 2001WO-US006520.
PR   22-MAR-2001; 2001US-00816744.
PR   22-MAR-2001; 2001US-00816920.
PR   22-MAR-2001; 2001WO-US009552.
PR   10-MAY-2001; 2001US-00854208.
PR   10-MAY-2001; 2001US-00854280.
PR   25-MAY-2001; 2001WO-US017092.
PR   01-JUN-2001; 2001US-00872035.
PR   01-JUN-2001; 2001WO-US017800.
PR   05-JUN-2001; 2001US-00874503.
PR   14-JUN-2001; 2001US-00882636.
PR   19-JUN-2001; 2001US-00886342.
PR   20-JUN-2001; 2001WO-US019692.
PR   29-JUN-2001; 2001WO-US021066.
PR   09-JUL-2001; 2001WO-US021735.
PR   30-JUL-2001; 2001US-00918585.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;

  Query Match             100.0%;  Score 1505;  DB 6;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 15
ABO33793
ID   ABO33793 standard; protein; 273 AA.
XX
AC   ABO33793;
XX
DT   17-SEP-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO1330.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   antiarthritic; osteopathic; gene therapy; TNF-Agonist-Alpha;
KW   chondrocyte stimulator; pericyte stimulator; fibroblast modulator;
KW   pharmaceutical; diagnostic; biosensor; bioreactor; tumour; lung tumour;
```

```
PR   01-DEC-2000; 2000WO-US032678.
PR   20-DEC-2000; 2000US-00747259.
PR   20-DEC-2000; 2000WO-US034956.
PR   28-FEB-2001; 2001WO-US006520.
PR   22-MAR-2001; 2001US-00816744.
PR   22-MAR-2001; 2001US-00816920.
PR   22-MAR-2001; 2001WO-US009552.
PR   10-MAY-2001; 2001US-00854208.
PR   10-MAY-2001; 2001US-00854280.
PR   25-MAY-2001; 2001WO-US017092.
PR   01-JUN-2001; 2001US-00872035.
PR   01-JUN-2001; 2001WO-US017800.
PR   05-JUN-2001; 2001US-00874503.
PR   14-JUN-2001; 2001US-00882636.
PR   19-JUN-2001; 2001US-00886342.
PR   20-JUN-2001; 2001WO-US019692.
PR   29-JUN-2001; 2001WO-US021066.
PR   09-JUL-2001; 2001WO-US021735.
PR   30-JUL-2001; 2001US-00918585.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;

  Query Match             100.0%;  Score 1505;  DB 6;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 14
ABO25272
ID   ABO25272 standard; protein; 273 AA.
XX
AC   ABO25272;
XX
DT   09-SEP-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO1330.
XX
KW   Human; secreted and transmembrane protein; PRO; virucide; gene therapy;
KW   cell death; growth induction cascade; blood coagulation cascade;
KW   viral infection.
XX
OS   Homo sapiens.
XX
PN   US2003050239-A1.
XX
PD   13-MAR-2003.
XX
PF   15-OCT-2001; 2001US-00978191.
XX
PR   17-OCT-1997;   97US-0062250P.
```

```
PR   03-NOV-1997;   97US-0064249P.
PR   13-NOV-1997;   97US-0065311P.
PR   21-NOV-1997;   97US-0066364P.
PR   10-MAR-1998;   98US-0077450P.
PR   11-MAR-1998;   98US-0077632P.
PR   11-MAR-1998;   98US-0077641P.
PR   11-MAR-1998;   98US-0077649P.
PR   12-MAR-1998;   98US-0077791P.
PR   13-MAR-1998;   98US-0078004P.
PR   17-MAR-1998;   98US-00040220.
PR   20-MAR-1998;   98US-0078886P.
PR   20-MAR-1998;   98US-0078910P.
PR   20-MAR-1998;   98US-0078936P.
PR   20-MAR-1998;   98US-0078939P.
PR   25-MAR-1998;   98US-0079294P.
PR   26-MAR-1998;   98US-0079656P.
PR   27-MAR-1998;   98US-0079663P.
PR   27-MAR-1998;   98US-0079664P.
PR   27-MAR-1998;   98US-0079689P.
PR   27-MAR-1998;   98US-0079728P.
PR   27-MAR-1998;   98US-0079786P.
PR   30-MAR-1998;   98US-0079920P.
PR   30-MAR-1998;   98US-0079923P.
PR   31-MAR-1998;   98US-0080105P.
PR   31-MAR-1998;   98US-0080107P.
PR   31-MAR-1998;   98US-0080165P.
PR   31-MAR-1998;   98US-0080194P.
PR   01-APR-1998;   98US-0080327P.
PR   01-APR-1998;   98US-0080328P.
PR   01-APR-1998;   98US-0080333P.
PR   01-APR-1998;   98US-0080334P.
PR   08-APR-1998;   98US-0081049P.
PR   08-APR-1998;   98US-0081070P.
PR   08-APR-1998;   98US-0081071P.
PR   09-APR-1998;   98US-0081195P.
PR   09-APR-1998;   98US-0081203P.
PR   09-APR-1998;   98US-0081229P.
PR   15-APR-1998;   98US-0081817P.
PR   15-APR-1998;   98US-0081819P.
PR   15-APR-1998;   98US-0081838P.
PR   15-APR-1998;   98US-0081952P.
PR   15-APR-1998;   98US-0081955P.
PR   21-APR-1998;   98US-0082568P.
PR   21-APR-1998;   98US-0082569P.
PR   22-APR-1998;   98US-0082700P.
PR   22-APR-1998;   98US-0082704P.
PR   22-APR-1998;   98US-0082797P.
PR   22-APR-1998;   98US-0082804P.
PR   23-APR-1998;   98US-0082796P.
PR   27-APR-1998;   98US-0083336P.
PR   28-APR-1998;   98US-0083322P.
PR   29-APR-1998;   98US-0083392P.
PR   29-APR-1998;   98US-0083495P.
PR   29-APR-1998;   98US-0083496P.
PR   29-APR-1998;   98US-0083499P.
PR   29-APR-1998;   98US-0083500P.
PR   29-APR-1998;   98US-0083545P.
PR   29-APR-1998;   98US-0083554P.
PR   29-APR-1998;   98US-0083558P.
PR   29-APR-1998;   98US-0083559P.
PR   30-APR-1998;   98US-0083742P.
PR   05-MAY-1998;   98US-0084366P.
PR   06-MAY-1998;   98US-0084414P.
PR   06-MAY-1998;   98US-0084441P.
PR   07-MAY-1998;   98US-0084598P.
PR   07-MAY-1998;   98US-0084600P.
PR   07-MAY-1998;   98US-0084627P.
PR   07-MAY-1998;   98US-0084637P.
PR   07-MAY-1998;   98US-0084639P.
PR   07-MAY-1998;   98US-0084640P.
PR   07-MAY-1998;   98US-0084643P.
PR   13-MAY-1998;   98US-0085323P.
PR   13-MAY-1998;   98US-0085338P.
```

```
PR   17-MAR-1998;   98US-00040220.
PR   20-MAR-1998;   98US-0078886P.
PR   20-MAR-1998;   98US-0078910P.
PR   20-MAR-1998;   98US-0078936P.
PR   20-MAR-1998;   98US-0078939P.
PR   25-MAR-1998;   98US-0079294P.
PR   26-MAR-1998;   98US-0079656P.
PR   27-MAR-1998;   98US-0079663P.
PR   27-MAR-1998;   98US-0079664P.
PR   27-MAR-1998;   98US-0079689P.
PR   27-MAR-1998;   98US-0079728P.
PR   27-MAR-1998;   98US-0079786P.
PR   30-MAR-1998;   98US-0079920P.
PR   30-MAR-1998;   98US-0079923P.
PR   31-MAR-1998;   98US-0080105P.
PR   31-MAR-1998;   98US-0080107P.
PR   31-MAR-1998;   98US-0080165P.
PR   31-MAR-1998;   98US-0080194P.
PR   01-APR-1998;   98US-0080327P.
PR   01-APR-1998;   98US-0080328P.
PR   01-APR-1998;   98US-0080333P.
PR   01-APR-1998;   98US-0080334P.
PR   08-APR-1998;   98US-0081049P.
PR   08-APR-1998;   98US-0081070P.
PR   08-APR-1998;   98US-0081071P.
PR   09-APR-1998;   98US-0081195P.
PR   09-APR-1998;   98US-0081203P.
PR   09-APR-1998;   98US-0081229P.
PR   15-APR-1998;   98US-0081817P.
PR   15-APR-1998;   98US-0081819P.
PR   15-APR-1998;   98US-0081838P.
PR   15-APR-1998;   98US-0081952P.
PR   15-APR-1998;   98US-0081955P.
PR   21-APR-1998;   98US-0082568P.
PR   21-APR-1998;   98US-0082569P.
PR   22-APR-1998;   98US-0082700P.
PR   22-APR-1998;   98US-0082704P.
PR   22-APR-1998;   98US-0082797P.
PR   22-APR-1998;   98US-0082804P.
PR   23-APR-1998;   98US-0082796P.
PR   27-APR-1998;   98US-0083336P.
PR   28-APR-1998;   98US-0083322P.
PR   29-APR-1998;   98US-0083392P.
PR   29-APR-1998;   98US-0083495P.
PR   29-APR-1998;   98US-0083496P.
PR   29-APR-1998;   98US-0083499P.
PR   29-APR-1998;   98US-0083500P.
PR   29-APR-1998;   98US-0083545P.
PR   29-APR-1998;   98US-0083554P.
PR   29-APR-1998;   98US-0083558P.
PR   29-APR-1998;   98US-0083559P.
PR   30-APR-1998;   98US-0083742P.
PR   05-MAY-1998;   98US-0084366P.
PR   06-MAY-1998;   98US-0084414P.
PR   06-MAY-1998;   98US-0084441P.
PR   07-MAY-1998;   98US-0084598P.
PR   07-MAY-1998;   98US-0084600P.
PR   07-MAY-1998;   98US-0084627P.
PR   07-MAY-1998;   98US-0084637P.
PR   07-MAY-1998;   98US-0084639P.
PR   07-MAY-1998;   98US-0084640P.
PR   07-MAY-1998;   98US-0084643P.
PR   13-MAY-1998;   98US-0085323P.
PR   13-MAY-1998;   98US-0085338P.
PR   13-MAY-1998;   98US-0085339P.
PR   15-MAY-1998;   98US-0085573P.
PR   15-MAY-1998;   98US-0085579P.
PR   15-MAY-1998;   98US-0085580P.
PR   15-MAY-1998;   98US-0085582P.
PR   15-MAY-1998;   98US-0085689P.
PR   15-MAY-1998;   98US-0085697P.
PR   15-MAY-1998;   98US-0085700P.
PR   15-MAY-1998;   98US-0085704P.
```

```
PR   18-MAY-1998;   98US-0086023P.
PR   22-MAY-1998;   98US-0086392P.
PR   22-MAY-1998;   98US-0086414P.
PR   22-MAY-1998;   98US-0086430P.
PR   22-MAY-1998;   98US-0086486P.
PR   28-MAY-1998;   98US-0087098P.
PR   28-MAY-1998;   98US-0087106P.
PR   28-MAY-1998;   98US-0087208P.
PR   26-JUN-1998;   98US-00105413.
PR   26-JUN-1998;   98US-0090863P.
PR   26-JUN-1998;   98US-0091010P.
PR   01-JUL-1998;   98US-0091359P.
PR   30-JUL-1998;   98US-0094651P.
PR   11-SEP-1998;   98US-0100038P.
PR   07-OCT-1998;   98US-00168978.
PR   07-OCT-1998;   98WO-US021141.
PR   02-NOV-1998;   98US-00184216.
PR   06-NOV-1998;   98US-00187368.
PR   20-NOV-1998;   98US-0109304P.
PR   20-NOV-1998;   98WO-US024855.
PR   07-DEC-1998;   98US-00202054.
PR   22-DEC-1998;   98US-00218517.
PR   22-DEC-1998;   98US-0113296P.
PR   23-DEC-1998;   98US-0113621P.
PR   05-JAN-1999;   99WO-US000106.
PR   05-MAR-1999;   99US-00254465.
PR   08-MAR-1999;   99WO-US005028.
PR   10-MAR-1999;   99US-00265686.
PR   10-MAR-1999;   99WO-US005190.
PR   12-MAR-1999;   99US-00267213.
PR   12-MAR-1999;   99US-0123957P.
PR   29-MAR-1999;   99US-0126773P.
PR   12-APR-1999;   99US-00284291.
PR   21-APR-1999;   99US-0130232P.
PR   26-APR-1999;   99US-0131022P.
PR   28-APR-1999;   99US-0131445P.
PR   14-MAY-1999;   99US-00311832.
PR   14-MAY-1999;   99US-0134287P.
PR   14-MAY-1999;   99WO-US010733.
PR   02-JUN-1999;   99WO-US012252.
PR   16-JUN-1999;   99US-0139557P.
PR   23-JUN-1999;   99US-0141037P.
PR   07-JUL-1999;   99US-0142680P.
PR   26-JUL-1999;   99US-0145698P.
PR   28-JUL-1999;   99US-0146222P.
PR   25-AUG-1999;   99US-00380137.
PR   25-AUG-1999;   99US-00380138.
PR   25-AUG-1999;   99US-00380142.
PR   29-OCT-1999;   99US-0162506P.
PR   30-NOV-1999;   99WO-US028313.
PR   02-DEC-1999;   99WO-US028551.
PR   02-DEC-1999;   99WO-US028565.
PR   16-DEC-1999;   99WO-US030095.
PR   30-DEC-1999;   99WO-US031243.
PR   30-DEC-1999;   99WO-US031274.
PR   05-JAN-2000; 2000WO-US000219.
PR   06-JAN-2000; 2000WO-US000277.
PR   06-JAN-2000; 2000WO-US000376.
PR   11-FEB-2000; 2000WO-US003565.
PR   18-FEB-2000; 2000WO-US004341.
PR   24-FEB-2000; 2000WO-US005004.
PR   02-MAR-2000; 2000WO-US005841.
PR   10-MAR-2000; 2000WO-US006319.
PR   21-MAR-2000; 2000WO-US007532.
PR   30-MAR-2000; 2000WO-US008439.
PR   17-MAY-2000; 2000WO-US013705.
PR   22-MAY-2000; 2000WO-US014042.
PR   30-MAY-2000; 2000WO-US014941.
PR   02-JUN-2000; 2000WO-US015264.
PR   28-JUL-2000; 2000WO-US020710.
PR   24-AUG-2000; 2000WO-US023328.
PR   08-NOV-2000; 2000US-00709238.
PR   27-NOV-2000; 2000US-00723749.
```

```
CC   treating proliferative disorders, inflammatory disorders. TANGO-175, or
CC   WDNM-2 activity also include apoptotic disorders, rheumatoid arthritis,
CC   systemic lupus erythematosus, insulin-dependent diabetes mellitus, immune
CC   -related disorders, e.g., immunodeficiency disorders, viral disorders,
CC   cell growth disorders, e.g., cancers and inflammatory disorders and
CC   apoptotic disorders. The nucleic acids of the invention are used in gene
CC   therapy. The present sequence is human T125 protein
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 5;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 12
ABU80827
ID   ABU80827 standard; protein; 273 AA.
XX
AC   ABU80827;
XX
DT   23-JUN-2003  (first entry)
XX
DE   Human PRO polypeptide #89.
XX
KW   Human; PRO polypeptide; secreted and transmembrane protein;
KW   anti-PRO antibody; diagnostic assay; gene expression; tumour; cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2003036635-A1.
XX
PD   20-FEB-2003.
XX
PF   28-AUG-2002; 2002US-00230163.
XX
PR   25-JUL-2000; 2000US-0220638P.
PR   01-JUN-2001; 2001WO-US017800.
PR   29-JUN-2001; 2001WO-US021066.
PR   09-APR-2002; 2002US-00119480.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Desnoyers L,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi JC,  Gurney AL,  Smith V,  Stephan JF,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2003-342045/32.
DR   N-PSDB; ACA66929.
XX
PT   One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT   useful for the manufacture of a medicament for diagnosing or treating
PT   tumor.
XX
```

```
PS   Claim 11; Fig 178; 314pp; English.
XX
CC   The present invention relates to the isolation of novel human PRO
CC   polypeptides, and the polynucleotide sequences encoding them. The PRO
CC   polypeptides are secreted and transmembrane proteins. The PRO
CC   polypeptides and polynucleotides are useful for preparing a medicament
CC   useful in the diagnosis and treatment of tumours. Anti-PRO antibodies are
CC   useful in diagnostic assays for PRO, by detecting its expression in
CC   specific cells, tissues or serum, and for affinity purification of PRO
CC   from recombinant cell culture or natural sources. ABU80739-ABU80860
CC   represent the human PRO polypeptides of the invention. Note: The sequence
CC   data for this patent was obtained in electronic format directly from the
CC   USPTO web site at seqdata.uspto.gov/psipsDIDEntry.html
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 6;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 13
ABO25271
ID   ABO25271 standard; protein; 273 AA.
XX
AC   ABO25271;
XX
DT   09-SEP-2003  (first entry)
XX
DE   Novel human secreted and transmembrane protein PRO213-1.
XX
KW   Human; secreted and transmembrane protein; PRO; virucide; gene therapy;
KW   cell death; growth induction cascade; blood coagulation cascade;
KW   viral infection.
XX
OS   Homo sapiens.
XX
PN   US2003050239-A1.
XX
PD   13-MAR-2003.
XX
PF   15-OCT-2001; 2001US-00978191.
XX
PR   17-OCT-1997;   97US-0062250P.
PR   03-NOV-1997;   97US-0064249P.
PR   13-NOV-1997;   97US-0065311P.
PR   21-NOV-1997;   97US-0066364P.
PR   10-MAR-1998;   98US-0077450P.
PR   11-MAR-1998;   98US-0077632P.
PR   11-MAR-1998;   98US-0077641P.
PR   11-MAR-1998;   98US-0077649P.
PR   12-MAR-1998;   98US-0077791P.
PR   13-MAR-1998;   98US-0078004P.
```

```
PI   Grimaldi JC,  Gurney AL,  Smith V,  Stephan JF,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2002-172001/22.
DR   N-PSDB; ABK33624.
XX
PT   One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT   useful for treating a PRO related disorder and for diagnosing tumors such
PT   as lung cancer, colon cancer, breast tumor, prostate tumor, rectal tumor
PT   or liver tumor.
XX
PS   Claim 11; Fig 178; 359pp; English.
XX
CC   The invention relates to one hundred and twenty two nucleic acids
CC   encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
CC   encode human secreted proteins. The PRO nucleic acids, polypeptides,
CC   agonists and antagonists are useful for treating a PRO related disorder.
CC   The PRO polypeptides are useful for diagnosing tumours, especially lung
CC   cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
CC   liver tumour. The PRO polypeptides are useful for stimulating the
CC   proliferation of, or gene expression, in pericyte cells, for stimulating
CC   the proliferation or differentiation of chondrocyte cells, for
CC   stimulating the release of tumour necrosis factor-alpha from human blood,
CC   for stimulating or inhibiting the proliferation of normal human dermal
CC   fibroblast cells. The PRO polypeptide may also be used as molecular
CC   weight markers and for tissue typing. The PRO nucleic acids have
CC   applications in molecular biology, including use as hybridisation probes,
CC   and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO
CC   protein sequences of the invention
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 5;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 11
AAE21079
ID   AAE21079 standard; protein; 273 AA.
XX
AC   AAE21079;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Human T125 (TANGO-125) protein.
XX
KW   Human; haematopoiesis; clotting; kidney failure; wound healing; cancer;
KW   neoplasia; pancreatic disorder; pancreatitis; cerebrovascular disease;
KW   heart disorder; ischaemic heart disease; neuroprotective; vulnerary;
KW   cardiovascular disorder; ischaemic heart disease; immunosuppressive;
KW   glomerular disease; glomerulonephritis; uterine disorder; hyperplasia;
KW   fetal spleen; prostate disorder; inflammatory disease; Crohn's disease;
KW   proliferative disorder; gynaecological; haemostatic; antibacterial;
```

```
KW   systemic lupus erythematosus; immunodeficiency disorder; antiasthmatic;
KW   cytostatic; nephrotropic; antidiabetic; cerebroprotective; tranquilliser;
KW   hypotensive; tumour; injury; trauma; antianginal; vasotropic; antiulcer;
KW   apoptotic disorder; rheumatoid arthritis; cardiant; renal disorder;
KW   hepatotropic; antipsoriatic; antiallergic; dermatological; virucide;
KW   T125.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .22
FT                   /label= Signal_peptide
FT   Protein         23. .273
FT                   /note= "Human mature T125 protein"
FT   Domain          107. .134
FT                   /note= "Epidermal growth factor domain 1"
FT   Domain          141. .176
FT                   /note= "Epidermal growth factor domain 2"
XX
PN   US2002028508-A1.
XX
PD   07-MAR-2002.
XX
PF   21-FEB-2001; 2001US-00790264.
XX
PR   23-APR-1998;   98US-00065363.
PR   23-APR-1998;   98US-00065661.
PR   22-JUN-1998;   98US-00102705.
PR   29-JUL-1998;   98US-00124538.
PR   23-APR-1999;   99US-00298531.
PR   22-JUN-1999;   99US-00337930.
PR   29-JUL-1999;   99US-00363630.
XX
PA   (HOLT/) HOLTZMAN D A.
PA   (GOOD/) GOODEARL A D J.
PA   (MCCA/) MCCARTHY S A.
XX
PI   Holtzman DA,  Goodearl ADJ,  Mccarthy SA;
XX
DR   WPI; 2002-303420/34.
DR   N-PSDB; AAD33536.
XX
PT   Novel TANGO polypeptides and nucleic acid molecules useful as modulating
PT   agents in regulating cellular processes and for diagnosing and treating
PT   heart, liver, lung, kidney, inflammatory and cellular proliferative
PT   disorders.
XX
PS   Claim 51; Fig 4; 138pp; English.
XX
CC   The invention relates to nucleic acids encoding a variety of proteins
CC   human T139 (TANGO-139), T125 (TANGO-125), T110 (TANGO-110), murine T175
CC   (TANGO-175), human T175 or murine WDNM-2, having diagnostic, preventive,
CC   therapeutic and other uses. Polypeptide of the invention has the ability
CC   to inhibit a proteinase activity, to modulate cell-cell interactions,
CC   haematopoiesis and the ability to modulate clotting. Polypeptide and
CC   polynucleotide of the invention are useful for diagnosing and treating
CC   disorder characterised by their aberrant expression or activity. The
CC   antibodies are useful as modulating agents in regulating a variety of
CC   cellular processes e.g. cell proliferation and/or cell differentiation.
CC   TANGO-139 is useful for treating kidney defects such as kidney failure,
CC   TANGO-125 is useful in wound healing and for treating cancer, TANGO-110
CC   is useful for treating neoplasia, TANGO-177 or WDNM-2 is useful for
CC   treating cancer, are useful to treat pancreatic disorders, such as
CC   pancreatitis, cerebrovascular disease, and tumours, and injury or trauma
CC   to the brain. TANGO-125, 110, 175 molecules treat heart disorders, e.g.,
CC   ischaemic heart disease, cardiovascular disorders, such as ischaemic
CC   heart disease. TANGO-139, 125, 110 and 175 molecules are useful to treat
CC   renal (kidney) disorders, such as glomerular disease (e.g., acute and
CC   chronic glomerulonephritis), TANGO-175 is useful to treat uterine
CC   disorders, hyperplasia of the endometrium. TANGO-110 is useful to treat
CC   spleen, e.g., the fetal spleen, associated diseases and disorder. TANGO-
CC   125 treats prostate disorders, such as inflammatory diseases, Crohn's
CC   disease and tumours. TANGO-139, 125, 110, 175 or WDNM-2 are useful for
```

```
RESULT 9
AAB24043
ID   AAB24043 standard; protein; 273 AA.
XX
AC   AAB24043;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Human PRO1330 protein sequence SEQ ID NO:6.
XX
KW   Human; tumour; diagnosis; neoplastic disease; identification; cancer;
KW   tumourigenesis; detection; neoplastic cell growth; proliferation;
KW   cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
KW   immunological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200053754-A1.
XX
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US000277.
XX
PR   08-MAR-1999;   99WO-US005028.
PR   12-MAR-1999;   99US-0123957P.
PR   29-MAR-1999;   99US-0126773P.
PR   21-APR-1999;   99US-0130232P.
PR   28-APR-1999;   99US-0131445P.
PR   05-OCT-1999;   99WO-US023089.
PR   30-NOV-1999;   99WO-US028313.
PR   02-DEC-1999;   99WO-US028551.
PR   02-DEC-1999;   99WO-US028564.
PR   30-DEC-1999;   99WO-US031243.
PR   30-DEC-1999;   99WO-US031274.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Desauvage FJ,  Goddard A,  Gurney AL,  Klein RD,  Roy MA;
PI   Wood WI;
XX
DR   WPI; 2000-572269/53.
DR   N-PSDB; AAC58227.
XX
PT   New isolated antibody for use in compositions and methods for the
PT   diagnosis and treatment of neoplastic cell growth and proliferation in
PT   mammals, including humans, and in monitoring tumor treatment.
XX
PS   Claim 61; Fig 6; 195pp; English.
XX
CC   The present invention describes an isolated antibody (Ab) that binds to
CC   one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
CC   PRO237, PRO324, PRO351, PRO362, PRO615, PRO531, PRO538, PRO3664, PRO618,
CC   PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions and
CC   methods for the diagnosis and treatment of neoplastic cell growth and
CC   proliferation in mammals, including humans. Genes and polypeptides
CC   encoded by them, that are amplified in the genome of a tumour cell, can
CC   be identified and are useful targets for the treatment and prevention of
CC   certain cancers and may be used to monitor tumour treatment. Compounds
CC   that inhibit the expression or activity of the identified polypeptides
CC   can be identified and used as antagonists. Benign or malignant tumours,
CC   inflammatory disorders and immunological disorders can be treated.
CC   AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
CC   in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
CC   AAB24041 to AAB24056 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 10
AAU83680
ID   AAU83680 standard; protein; 273 AA.
XX
AC   AAU83680;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Human PRO protein, Seq ID No 178.
XX
KW   Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW   breast cancer; prostate tumour; rectal tumour; liver tumour;
KW   pericyte cell proliferation; chondrocyte cell proliferation;
KW   tumour necrosis factor-alpha.
XX
OS   Homo sapiens.
XX
PN   WO200208288-A2.
XX
PD   31-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-US021066.
XX
PR   20-JUL-2000; 2000US-0219556P.
PR   25-JUL-2000; 2000US-0220585P.
PR   25-JUL-2000; 2000US-0220605P.
PR   25-JUL-2000; 2000US-0220607P.
PR   25-JUL-2000; 2000US-0220624P.
PR   25-JUL-2000; 2000US-0220638P.
PR   25-JUL-2000; 2000US-0220664P.
PR   25-JUL-2000; 2000US-0220666P.
PR   26-JUL-2000; 2000US-0220893P.
PR   28-JUL-2000; 2000WO-US020710.
PR   01-AUG-2000; 2000US-0222425P.
PR   22-AUG-2000; 2000US-0227133P.
PR   23-AUG-2000; 2000WO-US023522.
PR   24-AUG-2000; 2000WO-US023328.
PR   10-NOV-2000; 2000WO-US030873.
PR   28-NOV-2000; 2000US-0253646P.
PR   01-DEC-2000; 2000WO-US032678.
PR   20-DEC-2000; 2000US-00747259.
PR   20-DEC-2000; 2000WO-US034956.
PR   28-FEB-2001; 2001WO-US006520.
PR   01-MAR-2001; 2001WO-US006666.
PR   22-MAR-2001; 2001US-00816744.
PR   10-MAY-2001; 2001US-00854208.
PR   10-MAY-2001; 2001US-00854280.
PR   25-MAY-2001; 2001WO-US017092.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Desnoyers L,  Gerritsen ME,  Goddard A,  Godowski PJ;
```

```
PR   14-MAY-1999;   99US-0134287P.
PR   02-DEC-1999;   99WO-US028565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Ferrara N,  Goddard A,  Gurney AL,  Hillan KJ;
PI   Williams PM,  Wood WI;
XX
DR   WPI; 2000-638138/61.
DR   N-PSDB; AAA75703.
XX
PT   A composition useful for treatment and diagnosis of a cardiovascular,
PT   endothelial or angiogenic disorder, especially cancer, comprises (an
PT   agonist or antagonist of) a PRO320, PRO938, PRO1031, PRO296, PRO213,
PT   PRO1330 or PRO1449 polypeptide.
XX
PS   Claim 67; Fig 12; 152pp; English.
XX
CC   The present sequence represents PRO1330, a human notch 4 homologue. The
CC   specification describes PRO320, PRO938, PRO1031, PRO296, PRO213, PRO1330
CC   and PRO1449 polypeptides. The polypeptides promoter or inhibit
CC   angiogenesis and cardiovascularisation in mammals. The polypeptides are
CC   used for the treatment and diagnosis of a cardiovascular, endothelial or
CC   angiogenic disorder, especially cancer. Disorders that can be diagnosed,
CC   treated or prevented by the polypeptides of the invention include trauma
CC   such as wounds, artherosclerosis, and cardiac hypertrophy
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273

RESULT 8
AAB24042
ID   AAB24042 standard; protein; 273 AA.
XX
AC   AAB24042;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Human PRO213 protein sequence SEQ ID NO:4.
XX
KW   Human; tumour; diagnosis; neoplastic disease; identification; cancer;
KW   tumourigenesis; detection; neoplastic cell growth; proliferation;
KW   cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
KW   immunological disorder.
XX
OS   Homo sapiens.
XX
PN   WO200053754-A1.
XX
```

```
PD   14-SEP-2000.
XX
PF   06-JAN-2000; 2000WO-US000277.
XX
PR   08-MAR-1999;   99WO-US005028.
PR   12-MAR-1999;   99US-0123957P.
PR   29-MAR-1999;   99US-0126773P.
PR   21-APR-1999;   99US-0130232P.
PR   28-APR-1999;   99US-0131445P.
PR   05-OCT-1999;   99WO-US023089.
PR   30-NOV-1999;   99WO-US028313.
PR   02-DEC-1999;   99WO-US028551.
PR   02-DEC-1999;   99WO-US028564.
PR   30-DEC-1999;   99WO-US031243.
PR   30-DEC-1999;   99WO-US031274.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Desauvage FJ,  Goddard A,  Gurney AL,  Klein RD,  Roy MA;
PI   Wood WI;
XX
DR   WPI; 2000-572269/53.
DR   N-PSDB; AAC58226.
XX
PT   New isolated antibody for use in compositions and methods for the
PT   diagnosis and treatment of neoplastic cell growth and proliferation in
PT   mammals, including humans, and in monitoring tumor treatment.
XX
PS   Claim 61; Fig 4; 195pp; English.
XX
CC   The present invention describes an isolated antibody (Ab) that binds to
CC   one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
CC   PRO237, PRO324, PRO351, PRO362, PRO615, PRO531, PRO538, PRO3664, PRO618,
CC   PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions and
CC   methods for the diagnosis and treatment of neoplastic cell growth and
CC   proliferation in mammals, including humans. Genes and polypeptides
CC   encoded by them, that are amplified in the genome of a tumour cell, can
CC   be identified and are useful targets for the treatment and prevention of
CC   certain cancers and may be used to monitor tumour treatment. Compounds
CC   that inhibit the expression or activity of the identified polypeptides
CC   can be identified and used as antagonists. Benign or malignant tumours,
CC   inflammatory disorders and immunological disorders can be treated.
CC   AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
CC   in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
CC   AAB24041 to AAB24056 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
```

```
FT                   /note= "N-myristoylation site"
FT   Modified-site   93. .97
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   97. .103
FT                   /note= "N-myristoylation site"
FT   Modified-site   100. .106
FT                   /note= "N-myristoylation site"
FT   Modified-site   103. .109
FT                   /note= "N-myristoylation site"
FT   Region          123. .135
FT                   /note= "EGF-like domain cysteine pattern signature"
FT   Region          130. .133
FT                   /note= "cell attachment sequence"
FT   Modified-site   152. .164
FT                   /note= "aspartic acid and asparagine hydroxylation site"
FT   Modified-site   157. .163
FT                   /note= "N-myristoylation site"
FT   Modified-site   191. .197
FT                   /note= "N-myristoylation site"
FT   Modified-site   265. .271
FT                   /note= "N-myristoylation site"
XX
PN   WO200053752-A2.
XX
PD   14-SEP-2000.
XX
PF   30-DEC-1999;   99WO-US031274.
XX
PR   08-MAR-1999;   99WO-US005028.
PR   21-APR-1999;   99US-0130232P.
PR   26-APR-1999;   99US-0131022P.
PR   28-APR-1999;   99US-0131445P.
PR   14-MAY-1999;   99US-0134287P.
PR   02-DEC-1999;   99WO-US028565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Ferrara N,  Goddard A,  Gurney AL,  Hillan KJ;
PI   Williams PM,  Wood WI;
XX
DR   WPI; 2000-638138/61.
DR   N-PSDB; AAA75702.
XX
PT   A composition useful for treatment and diagnosis of a cardiovascular,
PT   endothelial or angiogenic disorder, especially cancer, comprises (an
PT   agonist or antagonist of) a PRO320, PRO938, PRO1031, PRO296, PRO213,
PT   PRO1330 or PRO1449 polypeptide.
XX
PS   Claim 67; Fig 10; 152pp; English.
XX
CC   The present sequence represents PRO213, a human growth arrest-specific
CC   gene 6 protein homologue. The specification describes PRO320, PRO938,
CC   PRO1031, PRO296, PRO213, PRO1330 and PRO1449 polypeptides. The
CC   polypeptides promoter or inhibit angiogenesis and cardiovascularisation
CC   in mammals. The polypeptides are used for the treatment and diagnosis of
CC   a cardiovascular, endothelial or angiogenic disorder, especially cancer.
CC   Disorders that can be diagnosed, treated or prevented by the polypeptides
CC   of the invention include trauma such as wounds, artherosclerosis, and
CC   cardiac hypertrophy
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
                |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 7
AAB18674
ID   AAB18674 standard; protein; 273 AA.
XX
AC   AAB18674;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of a human PRO1330 polypeptide.
XX
KW   Notch 4 homologue; PRO320; PRO938; PRO1031; PRO296; PRO213; PRO1449;
KW   angiogenesis; cardiovascularisation; PRO1330; cardiovascular disorder;
KW   endothelial disorder; angiogenic disorder; cancer; trauma; wound;
KW   artherosclerosis; cardiac hypertrophy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .19
FT                   /note= "signal sequence"
FT   Modified-site   19. .25
FT                   /note= "N-myristoylation site"
FT   Modified-site   26. .30
FT                   /note= "amidation site"
FT   Modified-site   78. .84
FT                   /note= "N-myristoylation site"
FT   Modified-site   93. .97
FT                   /note= "cAMP- and cGMP-dependent protein kinase
FT                   phosphorylation site"
FT   Modified-site   97. .103
FT                   /note= "N-myristoylation site"
FT   Modified-site   100. .106
FT                   /note= "N-myristoylation site"
FT   Modified-site   103. .109
FT                   /note= "N-myristoylation site"
FT   Region          123. .135
FT                   /note= "EGF-like domain cysteine pattern signature"
FT   Region          130. .133
FT                   /note= "cell attachment sequence"
FT   Modified-site   152. .164
FT                   /note= "aspartic acid and asparagine hydroxylation site"
FT   Modified-site   157. .163
FT                   /note= "N-myristoylation site"
FT   Modified-site   191. .197
FT                   /note= "N-myristoylation site"
FT   Modified-site   265. .271
FT                   /note= "N-myristoylation site"
XX
PN   WO200053752-A2.
XX
PD   14-SEP-2000.
XX
PF   30-DEC-1999;   99WO-US031274.
XX
PR   08-MAR-1999;   99WO-US005028.
PR   21-APR-1999;   99US-0130232P.
PR   26-APR-1999;   99US-0131022P.
PR   28-APR-1999;   99US-0131445P.
```

```
  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 5
AAB44325
ID   AAB44325 standard; protein; 273 AA.
XX
AC   AAB44325;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO213-1 protein sequence SEQ ID NO:506.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US004341.
XX
PR   08-MAR-1999;   99WO-US005028.
PR   12-MAR-1999;   99US-0123957P.
PR   29-MAR-1999;   99US-0126773P.
PR   21-APR-1999;   99US-0130232P.
PR   28-APR-1999;   99US-0131445P.
PR   14-MAY-1999;   99US-0134287P.
PR   23-JUN-1999;   99US-0141037P.
PR   26-JUL-1999;   99US-0145698P.
PR   29-OCT-1999;   99US-0162506P.
PR   30-NOV-1999;   99WO-US028313.
PR   02-DEC-1999;   99WO-US028551.
PR   02-DEC-1999;   99WO-US028565.
PR   16-DEC-1999;   99WO-US030095.
PR   30-DEC-1999;   99WO-US031243.
PR   30-DEC-1999;   99WO-US031274.
PR   05-JAN-2000; 2000WO-US000219.
PR   06-JAN-2000; 2000WO-US000277.
PR   06-JAN-2000; 2000WO-US000376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA,  Shelton DL;
PI   Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
```

```
DR   WPI; 2000-611443/58.
DR   N-PSDB; AAC78585.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods, to
PT   target bioactive molecules to specific cells, and to modulate cellular
PT   activities.
XX
PS   Claim 12; Fig 213; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed sequence
CC   tag) sequences which encode secreted or transmembrane PRO polypeptides.
CC   The PRO polynucleotides and polypeptides have cytostatic activity. The
CC   polynucleotides and polypeptides can be used for detecting the presence
CC   of PRO polypeptides in samples, for linking bioactive molecules to cells
CC   and for modulating biological activities of cells, using the polypeptides
CC   for specific targeting. The polypeptide targeting can be used to kill the
CC   target cells, e.g. for the treatment of cancers. The polypeptide pairs
CC   provide specific targeting of bioactive molecules to cells. AAC78600 to
CC   AAC78987 represent PCR primers and probes used in the isolation of the
CC   PRO polynucleotide sequences
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 6
AAB18673
ID   AAB18673 standard; protein; 273 AA.
XX
AC   AAB18673;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Amino acid sequence of a human a PRO213 polypeptide.
XX
KW   Growth arrest-specific gene 6 protein homologue; PRO320; PRO938; PRO1031;
KW   PRO296; PRO213; PRO1449; angiogenesis; cardiovascularisation; PRO1330;
KW   cardiovascular disorder; endothelial disorder; angiogenic disorder;
KW   cancer; trauma; wound; artherosclerosis; cardiac hypertrophy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .19
FT                   /note= "signal sequence"
FT   Modified-site   19. .25
FT                   /note= "N-myristoylation site"
FT   Modified-site   26. .30
FT                   /note= "amidation site"
FT   Modified-site   78. .84
```

```
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= Signal_peptide
FT   Protein         23..274
FT                   /label= TANGO_125
FT   Domain          107..134
FT                   /note= "EGF-like domain 1"
FT   Domain          141..176
FT                   /note= "EGF-like domain 2"
XX
PN   WO9954437-A2.
XX
PD   28-OCT-1999.
XX
PF   23-APR-1999;   99WO-US008900.
XX
PR   23-APR-1998;   98US-00065363.
PR   23-APR-1999;   99US-00298531.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   Holtzman DA;
XX
DR   WPI; 2000-013240/01.
DR   N-PSDB; AAZ37131.
XX
PT   Novel polynucleotides and polypeptides used to modulate a variety of
PT   cellular processes.
XX
PS   Claim 1; Fig 1; 120pp; English.
XX
CC   This is the human TANGO 125 (T125) amino acid sequence. The T125 protein
CC   has two epidermal growth factor (EGF)-like domains at amino acids 107-134
CC   and 141-176 and is predicted to have a molecular weight of approximately
CC   30kD. T125 is predicted to have no transmembrane domains and appears to
CC   be a secreted protein. There are three alternatively spliced forms of the
CC   T125 gene: T125a, T125b and T125c (AAZ37132-Z37135). The sequences of all
CC   variants of T125 are used in the invention to create antibodies which
CC   selectively bind to T125. The T125 polypeptide is used to modulate a
CC   variety of cellular processes. It can be used to produce fusion proteins.
CC   The protein may also be used to produce antibodies, and to identify T125
CC   antagonists and agonists. The T125 polynucleotides, polypeptides,
CC   homologues and antibodies can be used in screening assays; predictive
CC   medicine; and methods of treatment of T125 associated disorders. The T125
CC   polynucleotides can be used to express the protein; to detect T125 mRNA;
CC   to detect genetic alterations in the T125 gene; in forensic biology; and
CC   as a source of primers and probes
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 3;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
```

```
RESULT 4
AAB44326
ID   AAB44326 standard; protein; 273 AA.
XX
AC   AAB44326;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO1330 protein sequence SEQ ID NO:508.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US004341.
XX
PR   08-MAR-1999;   99WO-US005028.
PR   12-MAR-1999;   99US-0123957P.
PR   29-MAR-1999;   99US-0126773P.
PR   21-APR-1999;   99US-0130232P.
PR   28-APR-1999;   99US-0131445P.
PR   14-MAY-1999;   99US-0134287P.
PR   23-JUN-1999;   99US-0141037P.
PR   26-JUL-1999;   99US-0145698P.
PR   29-OCT-1999;   99US-0162506P.
PR   30-NOV-1999;   99WO-US028313.
PR   02-DEC-1999;   99WO-US028551.
PR   02-DEC-1999;   99WO-US028565.
PR   16-DEC-1999;   99WO-US030095.
PR   30-DEC-1999;   99WO-US031243.
PR   30-DEC-1999;   99WO-US031274.
PR   05-JAN-2000; 2000WO-US000219.
PR   06-JAN-2000; 2000WO-US000277.
PR   06-JAN-2000; 2000WO-US000376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA,  Shelton DL;
PI   Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
DR   N-PSDB; AAC78586.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods, to
PT   target bioactive molecules to specific cells, and to modulate cellular
PT   activities.
XX
PS   Claim 12; Fig 215; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed sequence
CC   tag) sequences which encode secreted or transmembrane PRO polypeptides.
CC   The PRO polynucleotides and polypeptides have cytostatic activity. The
CC   polynucleotides and polypeptides can be used for detecting the presence
CC   of PRO polypeptides in samples, for linking bioactive molecules to cells
CC   and for modulating biological activities of cells, using the polypeptides
CC   for specific targeting. The polypeptide targeting can be used to kill the
CC   target cells, e.g. for the treatment of cancers. The polypeptide pairs
CC   provide specific targeting of bioactive molecules to cells. AAC78600 to
CC   AAC78987 represent PCR primers and probes used in the isolation of the
CC   PRO polynucleotide sequences
XX
SQ   Sequence 273 AA;
```

```
PR   20-MAR-1998;   98US-0078939P.
PR   25-MAR-1998;   98US-0079294P.
PR   26-MAR-1998;   98US-0079656P.
PR   27-MAR-1998;   98US-0079663P.
PR   27-MAR-1998;   98US-0079664P.
PR   27-MAR-1998;   98US-0079689P.
PR   27-MAR-1998;   98US-0079728P.
PR   27-MAR-1998;   98US-0079786P.
PR   30-MAR-1998;   98US-0079920P.
PR   30-MAR-1998;   98US-0079923P.
PR   31-MAR-1998;   98US-0080105P.
PR   31-MAR-1998;   98US-0080107P.
PR   31-MAR-1998;   98US-0080165P.
PR   31-MAR-1998;   98US-0080194P.
PR   01-APR-1998;   98US-0080327P.
PR   01-APR-1998;   98US-0080328P.
PR   01-APR-1998;   98US-0080333P.
PR   01-APR-1998;   98US-0080334P.
PR   08-APR-1998;   98US-0081049P.
PR   08-APR-1998;   98US-0081070P.
PR   08-APR-1998;   98US-0081071P.
PR   09-APR-1998;   98US-0081195P.
PR   09-APR-1998;   98US-0081203P.
PR   09-APR-1998;   98US-0081229P.
PR   15-APR-1998;   98US-0081817P.
PR   15-APR-1998;   98US-0081838P.
PR   15-APR-1998;   98US-0081952P.
PR   15-APR-1998;   98US-0081955P.
PR   21-APR-1998;   98US-0082568P.
PR   21-APR-1998;   98US-0082569P.
PR   22-APR-1998;   98US-0082700P.
PR   22-APR-1998;   98US-0082704P.
PR   22-APR-1998;   98US-0082804P.
PR   23-APR-1998;   98US-0082767P.
PR   23-APR-1998;   98US-0082796P.
PR   27-APR-1998;   98US-0083336P.
PR   28-APR-1998;   98US-0083322P.
PR   29-APR-1998;   98US-0083392P.
PR   29-APR-1998;   98US-0083495P.
PR   29-APR-1998;   98US-0083496P.
PR   29-APR-1998;   98US-0083499P.
PR   29-APR-1998;   98US-0083500P.
PR   29-APR-1998;   98US-0083545P.
PR   29-APR-1998;   98US-0083554P.
PR   29-APR-1998;   98US-0083558P.
PR   29-APR-1998;   98US-0083559P.
PR   30-APR-1998;   98US-0083742P.
PR   05-MAY-1998;   98US-0084366P.
PR   06-MAY-1998;   98US-0084414P.
PR   06-MAY-1998;   98US-0084441P.
PR   07-MAY-1998;   98US-0084598P.
PR   07-MAY-1998;   98US-0084600P.
PR   07-MAY-1998;   98US-0084627P.
PR   07-MAY-1998;   98US-0084637P.
PR   07-MAY-1998;   98US-0084639P.
PR   07-MAY-1998;   98US-0084640P.
PR   07-MAY-1998;   98US-0084643P.
PR   13-MAY-1998;   98US-0085323P.
PR   13-MAY-1998;   98US-0085338P.
PR   13-MAY-1998;   98US-0085339P.
PR   15-MAY-1998;   98US-0085573P.
PR   15-MAY-1998;   98US-0085579P.
PR   15-MAY-1998;   98US-0085580P.
PR   15-MAY-1998;   98US-0085582P.
PR   15-MAY-1998;   98US-0085689P.
PR   15-MAY-1998;   98US-0085697P.
PR   15-MAY-1998;   98US-0085700P.
PR   15-MAY-1998;   98US-0085704P.
PR   18-MAY-1998;   98US-0086023P.
PR   22-MAY-1998;   98US-0086392P.
PR   22-MAY-1998;   98US-0086414P.
PR   22-MAY-1998;   98US-0086430P.
PR   22-MAY-1998;   98US-0086486P.
```

```
PR   28-MAY-1998;   98US-0087098P.
PR   28-MAY-1998;   98US-0087106P.
PR   28-MAY-1998;   98US-0087208P.
PR   30-JUL-1998;   98US-0094651P.
PR   11-SEP-1998;   98US-0100038P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wood WI,  Goddard A,  Gurney A,  Yuan J,  Baker KP,  Chen J;
XX
DR   WPI; 1999-551358/46.
DR   N-PSDB; AAZ34312.
XX
PT   New secreted and transmembrane polypeptides and their polynucleotides,
PT   useful for treating blood coagulation disorders, cancers and cellular
PT   adhesion disorders.
XX
PS   Claim 12; Fig 215; 530pp; English.
XX
CC   The present invention describes secreted and transmembrane polypeptides
CC   and their polynucleotides. The nucleotide sequences are useful as sources
CC   of probes, primers, for chromosome mapping, and for generation of
CC   antisense sequences. They can also be used to create transgenic animals.
CC   The proteins can be used to treat a variety of diseases and disorders,
CC   depending on their function. Diseases that may be treated include blood
CC   coagulation disorders, cancers and cellular adhesion disorders. They may
CC   also be used to raise antibodies. AAZ33891 to AAZ34338, and AAY41685 to
CC   AAY41774 represent polynucleotide and polypeptide sequence given in the
CC   exemplification of the present invention
XX
SQ   Sequence 273 AA;

  Query Match            100.0%;  Score 1505;  DB 2;  Length 273;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 3
AAY52137
ID   AAY52137 standard; protein; 273 AA.
XX
AC   AAY52137;
XX
DT   01-FEB-2000  (first entry)
XX
DE   Human TANGO 125 (T125) amino acid sequence.
XX
KW   TANGO 125; T125; alternative splice variant; EGF domain; antibody;
KW   secreted protein; agonist; antagonist; predictive medicine; treatment;
KW   forensic biology.
XX
OS   Homo sapiens.
XX
```

```
PR   01-APR-1998;   98US-0080333P.
PR   01-APR-1998;   98US-0080334P.
PR   08-APR-1998;   98US-0081049P.
PR   08-APR-1998;   98US-0081070P.
PR   08-APR-1998;   98US-0081071P.
PR   09-APR-1998;   98US-0081195P.
PR   09-APR-1998;   98US-0081203P.
PR   09-APR-1998;   98US-0081229P.
PR   15-APR-1998;   98US-0081817P.
PR   15-APR-1998;   98US-0081838P.
PR   15-APR-1998;   98US-0081952P.
PR   15-APR-1998;   98US-0081955P.
PR   21-APR-1998;   98US-0082568P.
PR   21-APR-1998;   98US-0082569P.
PR   22-APR-1998;   98US-0082700P.
PR   22-APR-1998;   98US-0082704P.
PR   22-APR-1998;   98US-0082804P.
PR   23-APR-1998;   98US-0082767P.
PR   23-APR-1998;   98US-0082796P.
PR   27-APR-1998;   98US-0083336P.
PR   28-APR-1998;   98US-0083322P.
PR   29-APR-1998;   98US-0083392P.
PR   29-APR-1998;   98US-0083495P.
PR   29-APR-1998;   98US-0083496P.
PR   29-APR-1998;   98US-0083499P.
PR   29-APR-1998;   98US-0083500P.
PR   29-APR-1998;   98US-0083545P.
PR   29-APR-1998;   98US-0083554P.
PR   29-APR-1998;   98US-0083558P.
PR   29-APR-1998;   98US-0083559P.
PR   30-APR-1998;   98US-0083742P.
PR   05-MAY-1998;   98US-0084366P.
PR   06-MAY-1998;   98US-0084414P.
PR   06-MAY-1998;   98US-0084441P.
PR   07-MAY-1998;   98US-0084598P.
PR   07-MAY-1998;   98US-0084600P.
PR   07-MAY-1998;   98US-0084627P.
PR   07-MAY-1998;   98US-0084637P.
PR   07-MAY-1998;   98US-0084639P.
PR   07-MAY-1998;   98US-0084640P.
PR   07-MAY-1998;   98US-0084643P.
PR   13-MAY-1998;   98US-0085323P.
PR   13-MAY-1998;   98US-0085338P.
PR   13-MAY-1998;   98US-0085339P.
PR   15-MAY-1998;   98US-0085573P.
PR   15-MAY-1998;   98US-0085579P.
PR   15-MAY-1998;   98US-0085580P.
PR   15-MAY-1998;   98US-0085582P.
PR   15-MAY-1998;   98US-0085689P.
PR   15-MAY-1998;   98US-0085697P.
PR   15-MAY-1998;   98US-0085700P.
PR   15-MAY-1998;   98US-0085704P.
PR   18-MAY-1998;   98US-0086023P.
PR   22-MAY-1998;   98US-0086392P.
PR   22-MAY-1998;   98US-0086414P.
PR   22-MAY-1998;   98US-0086430P.
PR   22-MAY-1998;   98US-0086486P.
PR   28-MAY-1998;   98US-0087098P.
PR   28-MAY-1998;   98US-0087106P.
PR   28-MAY-1998;   98US-0087208P.
PR   30-JUL-1998;   98US-0094651P.
PR   11-SEP-1998;   98US-0100038P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wood WI,  Goddard A,  Gurney A,  Yuan J,  Baker KP,  Chen J;
XX
DR   WPI; 1999-551358/46.
DR   N-PSDB; AAZ34311.
XX
PT   New secreted and transmembrane polypeptides and their polynucleotides,
PT   useful for treating blood coagulation disorders, cancers and cellular
PT   adhesion disorders.
XX
PS   Claim 12; Fig 213; 530pp; English.
XX
CC   The present invention describes secreted and transmembrane polypeptides
CC   and their polynucleotides. The nucleotide sequences are useful as sources
CC   of probes, primers, for chromosome mapping, and for generation of
CC   antisense sequences. They can also be used to create transgenic animals.
CC   The proteins can be used to treat a variety of diseases and disorders,
CC   depending on their function. Diseases that may be treated include blood
CC   coagulation disorders, cancers and cellular adhesion disorders. They may
CC   also be used to raise antibodies. AAZ33891 to AAZ34338, and AAY41685 to
CC   AAY41774 represent polynucleotide and polypeptide sequence given in the
CC   exemplification of the present invention
XX
SQ   Sequence 273 AA;

  Query Match             100.0%;  Score 1505;  DB 2;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-93;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 2
AAY41770
ID   AAY41770 standard; protein; 273 AA.
XX
AC   AAY41770;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human PRO1330 protein sequence.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein.
XX
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
PD   16-SEP-1999.
XX
PF   08-MAR-1999;   99WO-US005028.
XX
PR   10-MAR-1998;   98US-0077450P.
PR   11-MAR-1998;   98US-0077632P.
PR   11-MAR-1998;   98US-0077641P.
PR   11-MAR-1998;   98US-0077649P.
PR   12-MAR-1998;   98US-0077791P.
PR   13-MAR-1998;   98US-0078004P.
PR   17-MAR-1998;   98US-00040220.
PR   20-MAR-1998;   98US-0078886P.
PR   20-MAR-1998;   98US-0078910P.
PR   20-MAR-1998;   98US-0078936P.
```

```
                      GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         April 22, 2004, 12:41:14 ; Search time 61 Seconds
                                           (without alignments)
                                           1264.516 Million cell updates/sec

Title:          US-09-978-191A-506
Perfect score:  1505
Sequence:       1 MRGSQEVLLMWLLVLAVGGT..........SEQISFLEEQLGSCSCKKDS 273

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:      1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_29Jan04:*
               1:  geneseqp1980s:*
               2:  geneseqp1990s:*
               3:  geneseqp2000s:*
               4:  geneseqp2001s:*
               5:  geneseqp2002s:*
               6:  geneseqp2003as:*
               7:  geneseqp2003bs:*
               8:  geneseqp2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                     %
Result             Query
  No.    Score     Match Length DB  ID                    Description
 ------------------------------------------------------------------------
    1     1505     100.0    273  2  AAY41769              Aay41769 Human PRO
    2     1505     100.0    273  2  AAY41770              Aay41770 Human PRO
    3     1505     100.0    273  3  AAY52137              Aay52137 Human TAN
    4     1505     100.0    273  3  AAB44326              Aab44326 Human PRO
    5     1505     100.0    273  3  AAB44325              Aab44325 Human PRO
    6     1505     100.0    273  3  AAB18673              Aab18673 Amino aci
    7     1505     100.0    273  3  AAB18674              Aab18674 Amino aci
    8     1505     100.0    273  3  AAB24042              Aab24042 Human PRO
    9     1505     100.0    273  3  AAB24043              Aab24043 Human PRO
   10     1505     100.0    273  5  AAU83680              Aau83680 Human PRO
   11     1505     100.0    273  5  AAE21079              Aae21079 Human T12
   12     1505     100.0    273  6  ABU80827              Abu80827 Human PRO
   13     1505     100.0    273  6  ABO25271              Abo25271 Novel hum
   14     1505     100.0    273  6  ABO25272              Abo25272 Novel hum
   15     1505     100.0    273  6  ABO33793              Abo33793 Novel hum
   16     1505     100.0    273  6  ABU72277              Abu72277 Novel hum
   17     1505     100.0    273  6  ABU72278              Abu72278 Novel hum
   18     1505     100.0    273  6  ABU84957              Abu84957 Human sec
   19     1505     100.0    273  6  ABU84958              Abu84958 Human sec
   20     1505     100.0    273  6  ABU61156              Abu61156 Human PRO
   21     1505     100.0    273  6  ABU61155              Abu61155 Human PRO
   22     1505     100.0    273  6  ABU80424              Abu80424 Human sec
   23     1505     100.0    273  6  ABU80425              Abu80425 Human sec
   24     1505     100.0    273  6  ABU82136              Abu82136 Novel hum
   25     1505     100.0    273  6  ABJ72316              Abj72316 Human PRO
```

```
   26     1505     100.0    273  6  ADA25047              Ada25047 Novel hum
   27     1505     100.0    273  6  ADA25045              Ada25045 Novel hum
   28     1505     100.0    273  6  ABO19727              Abo19727 Novel hum
   29     1505     100.0    273  6  ABO19726              Abo19726 Novel hum
   30     1505     100.0    273  6  ADA12706              Ada12706 Human sec
   31     1505     100.0    273  6  ADA12708              Ada12708 Human sec
   32     1505     100.0    273  6  ABJ72444              Abj72444 Human PRO
   33     1505     100.0    273  6  ABO34339              Abo34339 Human sec
   34     1505     100.0    273  6  ABO19617              Abo19617 Novel hum
   35     1505     100.0    273  6  ABO19618              Abo19618 Novel hum
   36     1505     100.0    273  7  ABJ72146              Abj72146 Human mem
   37     1505     100.0    273  7  ADB83668              Adb83668 Novel hum
   38     1505     100.0    273  7  ADB80774              Adb80774 Novel hum
   39     1505     100.0    273  7  ADB73315              Adb73315 Novel hum
   40     1505     100.0    273  7  ADB78397              Adb78397 Novel hum
   41     1505     100.0    273  7  ADB85045              Adb85045 Human PRO
   42     1505     100.0    273  7  ADB78151              Adb78151 Novel hum
   43     1505     100.0    273  7  ADB74014              Adb74014 Human PRO
   44     1505     100.0    273  7  ADB74012              Adb74012 Human PRO
   45     1505     100.0    273  7  ADB87217              Adb87217 Human PRO

                                ALIGNMENTS


RESULT 1
AAY41769
ID   AAY41769 standard; protein; 273 AA.
XX
AC   AAY41769;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human PRO213-1 protein sequence.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein.
XX
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
PD   16-SEP-1999.
XX
PF   08-MAR-1999;   99WO-US005028.
XX
PR   10-MAR-1998;   98US-0077450P.
PR   11-MAR-1998;   98US-0077632P.
PR   11-MAR-1998;   98US-0077641P.
PR   11-MAR-1998;   98US-0077649P.
PR   12-MAR-1998;   98US-0077791P.
PR   13-MAR-1998;   98US-0078004P.
PR   17-MAR-1998;   98US-00040220.
PR   20-MAR-1998;   98US-0078886P.
PR   20-MAR-1998;   98US-0078910P.
PR   20-MAR-1998;   98US-0078936P.
PR   20-MAR-1998;   98US-0078939P.
PR   25-MAR-1998;   98US-0079294P.
PR   26-MAR-1998;   98US-0079656P.
PR   27-MAR-1998;   98US-0079663P.
PR   27-MAR-1998;   98US-0079664P.
PR   27-MAR-1998;   98US-0079689P.
PR   27-MAR-1998;   98US-0079728P.
PR   27-MAR-1998;   98US-0079786P.
PR   30-MAR-1998;   98US-0079920P.
PR   30-MAR-1998;   98US-0079923P.
PR   31-MAR-1998;   98US-0080105P.
PR   31-MAR-1998;   98US-0080107P.
PR   31-MAR-1998;   98US-0080165P.
PR   31-MAR-1998;   98US-0080194P.
PR   01-APR-1998;   98US-0080327P.
PR   01-APR-1998;   98US-0080328P.
```

```
FT   DISULFID   1208   1221       BY SIMILARITY.
FT   DISULFID   1223   1236       BY SIMILARITY.
FT   DISULFID   1242   1254       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1265   1278       BY SIMILARITY.
FT   DISULFID   1284   1296       BY SIMILARITY.
FT   DISULFID   1291   1305       BY SIMILARITY.
FT   DISULFID   1307   1320       BY SIMILARITY.
FT   DISULFID   1326   1339       BY SIMILARITY.
FT   DISULFID   1333   1348       BY SIMILARITY.
FT   DISULFID   1350   1361       BY SIMILARITY.
FT   DISULFID   1367   1380       BY SIMILARITY.
FT   DISULFID   1374   1389       BY SIMILARITY.
FT   DISULFID   1391   1402       BY SIMILARITY.
FT   DISULFID   1408   1420       BY SIMILARITY.
FT   DISULFID   1415   1429       BY SIMILARITY.
FT   DISULFID   1431   1444       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1456   1470       BY SIMILARITY.
FT   DISULFID   1472   1485       BY SIMILARITY.
FT   DISULFID   1491   1502       BY SIMILARITY.
FT   DISULFID   1497   1511       BY SIMILARITY.
FT   DISULFID   1513   1526       BY SIMILARITY.
FT   DISULFID   1610   1622       BY SIMILARITY.
FT   DISULFID   1617   1631       BY SIMILARITY.
FT   DISULFID   1633   1646       BY SIMILARITY.
FT   DISULFID   1652   1663       BY SIMILARITY.
FT   DISULFID   1658   1672       BY SIMILARITY.

  Query Match             11.5%;  Score 172.5;  DB 1;  Length 2871;
  Best Local Similarity   22.7%;  Pred. No. 1.7e-05;
  Matches   59;  Conservative   17;  Mismatches   81;  Indels  103;  Gaps    8;

Qy          26 PGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR---ACSTYRTIYRTAYRRSPGLAPA 82
               ||   | |        : |  |   | : ||   :   :| : |:|     |   | : :
Db          75 PGGNQCIVPICRHSCGDGFCSR---PNMCTCPSGQISPSCGS-RSIQHCNIRCMNGGSCS 130

Qy          83 RPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQ------- 135
                    |  |:  |      ||  :|:  | ||| || | || |  |: |  |:
Db         131 DDHCLCQKGYIGTH-----CGQPVCESGCLNGGRCVAPNRCACTYGFTGPQCERDYRTGP 185

Qy         136 -----------------------------------------------------------S 136

Db         186 CFTVVSNQMCQGQLSGIVCTKTLCCATVGRAWGHPCEMCPAQPHPCRRGFIPNIRTGACQ 245

Qy         137 DVDECSARRGGCP-QRCINTAGSYWCQCWEGHSLSADGTLC-----------VPKGG--- 181
               ||||| |  | |    |||| ||: |:|  ||  :     |           |  ||
Db         246 DVDECQAIPGMCQGGNCINTVGSFECKCPAGHKFNEVSQKCEDIDECSTIPGVCDGGECT 305

Qy         182 ----------PPRVAPNPTG 191
                         ||    :| |
Db         306 NTVSSYFCKCPPGFYTSPDG 325

Search completed: April 22, 2004, 12:45:50
Job time : 19 secs
```

```
CC       long-term force bearing structural support.
CC   -!- PTM: Forms intermolecular disulfide bonds either with other
CC       fibrillin-1 molecules or with other components of the microfibrils
CC       (By similarity).
CC   -!- SIMILARITY: Contains 47 EGF-like domains.
CC   -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L29454; AAA56840.1; -.
DR   EMBL; U22493; AAA64217.1; -.
DR   PIR; A55624; A55624.
DR   HSSP; P35555; 1APJ.
DR   MGD; MGI:95489; Fbn1.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 45.
DR   Pfam; PF00683; TB; 9.
DR   SMART; SM00179; EGF_CA; 42.
DR   PROSITE; PS00010; ASX_HYDROXYL; 43.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 38.
DR   PROSITE; PS50026; EGF_3; 45.
DR   PROSITE; PS01187; EGF_CA; 43.
KW   Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW   Repeat; Signal; Multigene family.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28   2871       FIBRILLIN 1.
FT   DOMAIN       81    112       EGF-LIKE 1.
FT   DOMAIN      115    146       EGF-LIKE 2.
FT   DOMAIN      147    178       EGF-LIKE 3.
FT   DOMAIN      246    287       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      288    329       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN      330    401       TGFBP 1.
FT   DOMAIN      402    446       PRO-RICH.
FT   DOMAIN      449    489       EGF-LIKE 6.
FT   DOMAIN      490    529       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN      530    571       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN      572    612       EGF-LIKE 9, CALCIUM-BINDING.
FT   DOMAIN      613    653       EGF-LIKE 10, CALCIUM-BINDING.
FT   DOMAIN      656    721       TGFBP 2.
FT   DOMAIN      723    764       EGF-LIKE 11, CALCIUM-BINDING.
FT   DOMAIN      765    806       EGF-LIKE 12, CALCIUM-BINDING.
FT   DOMAIN      807    846       EGF-LIKE 13, CALCIUM-BINDING.
FT   DOMAIN      910    951       EGF-LIKE 14, CALCIUM-BINDING.
FT   DOMAIN      952   1018       TGFBP 3.
FT   DOMAIN     1028   1069       EGF-LIKE 15, CALCIUM-BINDING.
FT   DOMAIN     1070   1112       EGF-LIKE 16, CALCIUM-BINDING.
FT   DOMAIN     1113   1154       EGF-LIKE 17, CALCIUM-BINDING.
FT   DOMAIN     1155   1196       EGF-LIKE 18, CALCIUM-BINDING.
FT   DOMAIN     1197   1237       EGF-LIKE 19, CALCIUM-BINDING.
FT   DOMAIN     1238   1279       EGF-LIKE 20, CALCIUM-BINDING.
FT   DOMAIN     1280   1321       EGF-LIKE 21, CALCIUM-BINDING.
FT   DOMAIN     1322   1362       EGF-LIKE 22, CALCIUM-BINDING.
FT   DOMAIN     1363   1403       EGF-LIKE 23, CALCIUM-BINDING.
FT   DOMAIN     1404   1445       EGF-LIKE 24, CALCIUM-BINDING.
FT   DOMAIN     1446   1486       EGF-LIKE 25, CALCIUM-BINDING.
FT   DOMAIN     1487   1527       EGF-LIKE 26, CALCIUM-BINDING.
FT   DOMAIN     1528   1599       TGFBP 4.
FT   DOMAIN     1606   1647       EGF-LIKE 27, CALCIUM-BINDING.
FT   DOMAIN     1648   1688       EGF-LIKE 28, CALCIUM-BINDING.
FT   DOMAIN     1689   1758       TGFBP 5.
FT   DOMAIN     1766   1807       EGF-LIKE 29, CALCIUM-BINDING.
FT   DOMAIN     1808   1848       EGF-LIKE 30, CALCIUM-BINDING.
FT   DOMAIN     1849   1890       EGF-LIKE 31, CALCIUM-BINDING.
```

```
FT   DOMAIN     1891   1929       EGF-LIKE 32, CALCIUM-BINDING.
FT   DOMAIN     1930   1972       EGF-LIKE 33, CALCIUM-BINDING.
FT   DOMAIN     1973   2012       EGF-LIKE 34, CALCIUM-BINDING.
FT   DOMAIN     2013   2054       EGF-LIKE 35, CALCIUM-BINDING.
FT   DOMAIN     2055   2121       TGFBP 6.
FT   DOMAIN     2127   2165       EGF-LIKE 36, CALCIUM-BINDING.
FT   DOMAIN     2166   2205       EGF-LIKE 37, CALCIUM-BINDING.
FT   DOMAIN     2206   2246       EGF-LIKE 38, CALCIUM-BINDING.
FT   DOMAIN     2247   2290       EGF-LIKE 39, CALCIUM-BINDING.
FT   DOMAIN     2291   2332       EGF-LIKE 40, CALCIUM-BINDING.
FT   DOMAIN     2333   2400       TGFBP 7.
FT   DOMAIN     2402   2443       EGF-LIKE 41, CALCIUM-BINDING.
FT   DOMAIN     2444   2484       EGF-LIKE 42, CALCIUM-BINDING.
FT   DOMAIN     2485   2523       EGF-LIKE 43, CALCIUM-BINDING.
FT   DOMAIN     2524   2566       EGF-LIKE 44, CALCIUM-BINDING.
FT   DOMAIN     2567   2606       EGF-LIKE 45, CALCIUM-BINDING.
FT   DOMAIN     2607   2647       EGF-LIKE 46, CALCIUM-BINDING.
FT   DOMAIN     2648   2687       EGF-LIKE 47, CALCIUM-BINDING.
FT   DISULFID     85     94       BY SIMILARITY.
FT   DISULFID     89    100       BY SIMILARITY.
FT   DISULFID    102    111       BY SIMILARITY.
FT   DISULFID    119    129       BY SIMILARITY.
FT   DISULFID    123    134       BY SIMILARITY.
FT   DISULFID    136    145       BY SIMILARITY.
FT   DISULFID    150    160       BY SIMILARITY.
FT   DISULFID    154    166       BY SIMILARITY.
FT   DISULFID    168    177       BY SIMILARITY.
FT   DISULFID    250    262       BY SIMILARITY.
FT   DISULFID    257    271       BY SIMILARITY.
FT   DISULFID    273    286       BY SIMILARITY.
FT   DISULFID    292    304       BY SIMILARITY.
FT   DISULFID    299    313       BY SIMILARITY.
FT   DISULFID    315    328       BY SIMILARITY.
FT   DISULFID    453    465       BY SIMILARITY.
FT   DISULFID    460    474       BY SIMILARITY.
FT   DISULFID    476    488       BY SIMILARITY.
FT   DISULFID    494    504       BY SIMILARITY.
FT   DISULFID    499    513       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    534    546       BY SIMILARITY.
FT   DISULFID    541    555       BY SIMILARITY.
FT   DISULFID    557    570       BY SIMILARITY.
FT   DISULFID    576    587       BY SIMILARITY.
FT   DISULFID    582    596       BY SIMILARITY.
FT   DISULFID    598    611       BY SIMILARITY.
FT   DISULFID    617    628       BY SIMILARITY.
FT   DISULFID    623    637       BY SIMILARITY.
FT   DISULFID    639    652       BY SIMILARITY.
FT   DISULFID    727    739       BY SIMILARITY.
FT   DISULFID    734    748       BY SIMILARITY.
FT   DISULFID    750    763       BY SIMILARITY.
FT   DISULFID    769    781       BY SIMILARITY.
FT   DISULFID    776    790       BY SIMILARITY.
FT   DISULFID    792    805       BY SIMILARITY.
FT   DISULFID    811    821       BY SIMILARITY.
FT   DISULFID    816    830       BY SIMILARITY.
FT   DISULFID    832    845       BY SIMILARITY.
FT   DISULFID    914    926       BY SIMILARITY.
FT   DISULFID    921    935       BY SIMILARITY.
FT   DISULFID    937    950       BY SIMILARITY.
FT   DISULFID   1032   1044       BY SIMILARITY.
FT   DISULFID   1039   1053       BY SIMILARITY.
FT   DISULFID   1055   1068       BY SIMILARITY.
FT   DISULFID   1074   1086       BY SIMILARITY.
FT   DISULFID   1081   1095       BY SIMILARITY.
FT   DISULFID   1097   1111       BY SIMILARITY.
FT   DISULFID   1117   1129       BY SIMILARITY.
FT   DISULFID   1124   1138       BY SIMILARITY.
FT   DISULFID   1140   1153       BY SIMILARITY.
FT   DISULFID   1159   1171       BY SIMILARITY.
FT   DISULFID   1166   1180       BY SIMILARITY.
FT   DISULFID   1182   1195       BY SIMILARITY.
FT   DISULFID   1201   1212       BY SIMILARITY.
```

```
DR   EMBL; X84819; CAA57709.1; JOINED.
DR   EMBL; X84820; CAA57709.1; JOINED.
DR   EMBL; X84821; CAA57709.1; JOINED.
DR   EMBL; X84822; CAA57709.1; JOINED.
DR   EMBL; X84823; CAA57709.1; JOINED.
DR   EMBL; X84824; CAA57709.1; JOINED.
DR   EMBL; X84825; CAA57709.1; JOINED.
DR   EMBL; X84826; CAA57709.1; JOINED.
DR   EMBL; X84827; CAA57709.1; JOINED.
DR   EMBL; X84828; CAA57709.1; JOINED.
DR   EMBL; X84829; CAA57709.1; JOINED.
DR   EMBL; X84830; CAA57709.1; JOINED.
DR   EMBL; X84831; CAA57709.1; JOINED.
DR   EMBL; X84832; CAA57709.1; JOINED.
DR   EMBL; X84833; CAA57709.1; JOINED.
DR   EMBL; X84834; CAA57709.1; JOINED.
DR   EMBL; X84835; CAA57709.1; JOINED.
DR   EMBL; X84836; CAA57709.1; JOINED.
DR   EMBL; X84837; CAA57709.1; JOINED.
DR   EMBL; M27445; AAA57261.1; -.
DR   PIR; A33322; MMHUND.
DR   PDB; 1NDX; 29-DEC-99.
DR   Genew; HGNC:7821; NID.
DR   MIM; 131390; -.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006605; G2F.
DR   InterPro; IPR009017; GFP_like.
DR   InterPro; IPR000033; Ldl_receptor_rep.
DR   InterPro; IPR003886; Nidogen_ext.
DR   InterPro; IPR000716; Thyroglobulin_1.
DR   Pfam; PF00008; EGF; 6.
DR   Pfam; PF00058; ldl_recept_b; 3.
DR   Pfam; PF00086; thyroglobulin_1; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00682; G2F; 1.
DR   SMART; SM00135; LY; 5.
DR   SMART; SM00539; NIDO; 1.
DR   SMART; SM00211; TY; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS50026; EGF_3; 5.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;
KW   Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion;
KW   3D-structure.
FT   SIGNAL        1     28
FT   CHAIN        29   1247       NIDOGEN.
FT   DOMAIN       29    669       I (LARGER GLOBULAR DOMAIN).
FT   DOMAIN      670    917       II (CYSTEINE-RICH).
FT   DOMAIN      918   1247       III (SMALLER GLOBULAR DOMAIN).
FT   DOMAIN      386    426       EGF-LIKE 1.
FT   DOMAIN      668    709       EGF-LIKE 2.
FT   DOMAIN      710    751       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      758    801       EGF-LIKE 4.
FT   DOMAIN      802    840       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      872    919       THYROGLOBULIN TYPE-I.
FT   DOMAIN      989   1030       LDL-RECEPTOR YWTD MOTIF 1.
FT   DOMAIN     1032   1073       LDL-RECEPTOR YWTD MOTIF 2.
FT   DOMAIN     1075   1118       LDL-RECEPTOR YWTD MOTIF 3.
FT   DOMAIN     1124   1163       LDL-RECEPTOR YWTD MOTIF 4.
FT   DOMAIN     1208   1244       EGF-LIKE 6.
FT   MOD_RES     289    289       SULFATION (POTENTIAL).
FT   MOD_RES     296    296       SULFATION (POTENTIAL).
FT   DISULFID    672    685       BY SIMILARITY.
FT   DISULFID    679    695       BY SIMILARITY.
FT   DISULFID    697    708       BY SIMILARITY.
FT   DISULFID    714    727       BY SIMILARITY.
FT   DISULFID    721    736       BY SIMILARITY.
FT   DISULFID    738    750       BY SIMILARITY.
```

```
FT   DISULFID    762    777       BY SIMILARITY.
FT   DISULFID    769    787       BY SIMILARITY.
FT   DISULFID    789    800       BY SIMILARITY.
FT   DISULFID    806    817       BY SIMILARITY.
FT   DISULFID    811    826       BY SIMILARITY.
FT   DISULFID    828    839       BY SIMILARITY.
FT   DISULFID   1212   1223       BY SIMILARITY.
FT   DISULFID   1219   1232       BY SIMILARITY.
FT   DISULFID   1234   1243       BY SIMILARITY.
FT   SITE        702    704       CELL ATTACHMENT SITE.
FT   CARBOHYD   1137   1137       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     33     34       EL -> SS (IN REF. 2).
FT   CONFLICT     37     42       FGPGQG -> SAPDR (IN REF. 2).
FT   CONFLICT   1115   1115       H -> T (IN REF. 3).
SQ   SEQUENCE   1247 AA;  136488 MW;  4681B5B3CEC1575B CRC64;

  Query Match             11.5%;  Score 173.5;  DB 1;  Length 1247;
  Best Local Similarity   25.8%;  Pred. No. 6e-06;
  Matches   80;  Conservative   23;  Mismatches   98;  Indels  109;  Gaps   18;

Qy          20 TEHAYRPGRRV-----CAVRAHG--------DPVSESFVQRVYQPFLTTCDGHRACSTYR 66
               |  | ||| |      |::   |        |  ||       ||  : |  |  |: :
Db         681 TNAACRPGPRTQFTCECSIGFRGDGRTCYDIDECSE-------QP--SVCGSHTICNNH- 730

Qy          67 TIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPP---CRNG--------- 114
                             |   |  |  |:: :    | | | : | |   |  |
Db         731 --------------PGTFRCECVEGYQFSD--EGTCVAVVDQRPINYCETGLHNCDIPQR 774

Qy         115 GSCVQPG----RCRCPAGWRGD--TCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHS 168
                 |:  |     | |  |: ||   || |||||   |      | || ||: |||  |:
Db         775 AQCIYTGGSSYTCSCLPGFSGDGQACQ-DVDECQPSRCHPDAFCYNTPGSFTCQCKPGY- 832

Qy         169 LSADGTLCVP------------------KGGPPRVAPNPTG------------------- 191
                  ||  |||                   |      | | |
Db         833 -QGDGFRCVPGEVEKTRCQHEREHILGAAGATDPQRPIPPGLFVPECDAHGHYAPTQCHG 891

Qy         192 -------VDSAMKEEVQRLQSRVDLLEEKLQLVLAPLH-SLASQALEHGLPDPGSLLVHS 243
                      ||   : ||:  ::|  :    |  |  |:|   |       || ||: |:
Db         892 STGYCWCVDRDGR-EVEGTRTRPGMTPPCLSTVAPPIHQGPAVPTAVIPLP-PGTHLL-- 947

Qy         244 FQQLGRIDSL 253
               | | |:|: |
Db         948 FAQTGKIERL 957


RESULT 15
FBN1_MOUSE
ID   FBN1_MOUSE     STANDARD;      PRT;  2871 AA.
AC   Q61554; Q60826;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fibrillin 1 precursor.
GN   FBN1 OR FBN-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95130561; PubMed=7829516;
RA   Yin W., Germiller J., Sanguineti C., Smiley E., Pangilinan T.,
RA   Pereira L., Ramirez F., Bonadio J.;
RT   "Primary structure and developmental expression of Fbn-1, the mouse
RT   fibrillin gene.";
RL   J. Biol. Chem. 270:1798-1806(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1; TISSUE=Kidney;
RA   Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Structural component of connective tissue microfibrils
CC       that binds calcium. Fibrillin-1-containing microfibrils provide
```

```
RA   Lieber T., Kidd S., Young M.W.;
RT   "kuzbanian-mediated cleavage of Drosophila Notch.";
RL   Genes Dev. 16:209-221(2002).
RN   [12]
RP   MUTANT MCD5.
RX   MEDLINE=21575956; PubMed=11719214;
RA   Ramain P., Khechumian K., Seugnet L., Arbogast N., Ackermann C.,
RA   Heitzler P.;
RT   "Novel Notch alleles reveal a Deltex-dependent pathway repressing
RT   neural fate.";
RL   Curr. Biol. 11:1729-1738(2001).
RN   [13]
RP   REVIEW.
RX   MEDLINE=22256570; PubMed=12369105;
RA   Portin P.;
RT   "General outlines of the molecular genetics of the Notch signalling
RT   pathway in Drosophila melanogaster: a review.";
RL   Hereditas 136:89-96(2002).
CC   -!- FUNCTION: Signaling protein, which regulates, with both positive
CC       and negative signals, the differentiation of at least central and
CC       peripheral nervous system and eye, wing disk, oogenesis, segmental
CC       appendages such as antennae and legs, and muscles, through lateral
CC       inhibition or induction. Functions as a receptor for membrane-
CC       bound ligands Delta and Serrate to regulate cell-fate
CC       determination. Upon ligand activation, and releasing from the cell
CC       membrane, the Notch intracellular domain (NICD) forms a
CC       transcriptional activator complex with Su(H) (Suppressor of
CC       hairless) and activates genes of the E(spl) complex. Essential for
CC       proper differentiation of ectoderm.
CC   -!- SUBUNIT: Interacts with Su(H) when activated. Interacts with Dx
CC       via its ANK repeats.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Upon activation and
CC       S3 cleavage, it is released from the cell membrane and enters into
CC       the nucleus in conjunction with Su(H).
CC   -!- PTM: Upon binding its ligands such as Delta or Serrate, it is
CC       cleaved (S2 cleavage) in its extracellular domain, close to the
CC       transmembrane domain. S2 cleavage is probably mediated by Kuz. It
CC       is then cleaved (S3 cleavage) downstream of its transmembrane
CC       domain, releasing it from the cell membrane. S3 cleavage requires
CC       Psn.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 36 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 6 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M16152; AAB59220.1; -.
DR   EMBL; M16153; AAB59220.1; JOINED.
DR   EMBL; M16149; AAB59220.1; JOINED.
DR   EMBL; M16150; AAB59220.1; JOINED.
DR   EMBL; M16151; AAB59220.1; JOINED.
DR   EMBL; K03508; AAA28725.1; -.
DR   EMBL; M13689; AAA28725.1; JOINED.
DR   EMBL; K03507; AAA28725.1; JOINED.
DR   EMBL; AE003426; AAF45848.2; -.
DR   EMBL; AL035436; CAB37610.1; -.
DR   EMBL; AL035395; CAB37610.1; JOINED.
DR   EMBL; M12175; AAA74496.1; -.
DR   EMBL; M16025; AAA28726.1; -.

  Query Match             11.6%;  Score 175;  DB 1;  Length 2703;
  Best Local Similarity   25.3%;  Pred. No. 1e-05;
  Matches   57;  Conservative   22;  Mismatches   62;  Indels   84;  Gaps   13;

Qy          18 GGTEHAYRPGRRVCAVRAH--GDPVSESFVQRVYQPFLTTCDGHRACSTYRT----IYRT 71
               |||      |:  ||  :|  ||                 |:    |:: |       :
Db          70 GGTCVTQLNGKTYCACDSHYVGD----------------YCEHRNPCNSMRCQNGGTCQV 113

Qy          72 AYRRSPGLAPARPRYAC-CP-GWKRT---SGLPGACGAAIC------------------- 107
                :|        ||  :| ||  |:  :     :| ||    |
Db         114 TFRN------GRPGISCKCPLGFDESLCEIAVPNACDHVTCLNGGTCQLKTLEEYTCACA 167

Qy         108 ---------------QPPCRNGGSCV-----QPGRCRCPAGWRGDTCQSDVDECSA---R 144
                                ||||| :|          | || |: ||||  |::|| :   :
Db         168 NGYTGERCETKNLCASSPCRNGATCTALAGSSSFTCSCPPGFTGDTCSYDIEECQSNPCK 227

Qy         145 RGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKGGPPRVAPNP 189
                ||    |:|| ||| | |  |::    |  |  |  |   :|:|
Db         228 YGG---TCVNTHGSYQCMCPTGYT----GKDCDTKYKP--CSPSP 263

RESULT 14
NIDO_HUMAN
ID   NIDO_HUMAN     STANDARD;      PRT;  1247 AA.
AC   P14543; Q14942;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Nidogen precursor (Entactin).
GN   NID.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90091745; PubMed=2574658;
RA   Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
RA   Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
RT   "Human nidogen: complete amino acid sequence and structural domains
RT   deduced from cDNAs, and evidence for polymorphism of the gene.";
RL   DNA 8:581-594(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96044428; PubMed=7557988;
RA   Zimmermann K., Hoischen S., Hafner M., Nischt R.;
RT   "Genomic sequences and structural organization of the human nidogen
RT   gene (NID).";
RL   Genomics 27:245-250(1995).
RN   [3]
RP   SEQUENCE OF 667-1247 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=89270475; PubMed=2471408;
RA   Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Passage E., Weil D.,
RA   Timpl R., Chu M.L., Uitto J.;
RT   "Human nidogen: cDNA cloning, cellular expression, and mapping of the
RT   gene to chromosome 1q43.";
RL   Am. J. Hum. Genet. 44:876-885(1989).
CC   -!- FUNCTION: Sulfated glycoprotein which is widely distributed in
CC       basement membranes and that is tightly associated with laminin.
CC       Also binds to collagen IV. It probably has a role in cell-
CC       extracellular matrix interactions.
CC   -!- SUBUNIT: Interacts with FBLN1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Basement membranes.
CC   -!- PTM: N- and O-glycosylated.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 thyroglobulin type-I domain.
CC   -!- SIMILARITY: Contains 5 LDL-receptor YWTD domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M30269; AAA59932.1; -.
DR   EMBL; X82245; CAA57709.1; -.
```

```
Qy        131 --GDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSA-DGTL------CVPKGG 181
                |:    |||||:  |  | | | || ||: | | ||: |:  |||       ||  ||
Db        377 GPGEGACQDVDECALGRSPCAQGCTNTDGSFHCSCEEGYVLAGEDGTQCQDVDECVGPGG 436

Qy        182 P 182
              |
Db        437 P 437


RESULT 13
NOTC_DROME
ID   NOTC_DROME     STANDARD;      PRT;  2703 AA.
AC   P07207; O97458; P04154; Q9W4T8;
DT   01-NOV-1986 (Rel. 03, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurogenic locus Notch protein precursor.
GN   N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R; TISSUE=Embryo;
RX   MEDLINE=86079539; PubMed=3935325;
RA   Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT   "Nucleotide sequence from the neurogenic locus notch implies a gene
RT   product that shares homology with proteins containing EGF-like
RT   repeats.";
RL   Cell 43:567-581(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S, and Oregon-R; TISSUE=Embryo;
RX   MEDLINE=87064624; PubMed=3097517;
RA   Kidd S., Kelley M.R., Young M.W.;
RT   "Sequence of the notch locus of Drosophila melanogaster: relationship
RT   of the encoded protein to mammalian clotting and growth factors.";
RL   Mol. Cell. Biol. 6:3094-3108(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

```
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=20196011; PubMed=10731137;
RA   Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA   Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA   Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA   Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA   Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA   Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,
RA   Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,
RA   Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA   McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA   Glover D.M.;
RT   "From sequence to chromosome: the tip of the X chromosome of D.
RT   melanogaster.";
RL   Science 287:2220-2222(2000).
RN   [5]
RP   SEQUENCE OF 2505-2611 FROM N.A.
RX   MEDLINE=85099329; PubMed=2981631;
RA   Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
RT   "opa: a novel family of transcribed repeats shared by the Notch locus
RT   and other developmentally regulated loci in D. melanogaster.";
RL   Cell 40:55-62(1985).
RN   [6]
RP   SEQUENCE OF 1-8 FROM N.A.
RX   MEDLINE=87257846; PubMed=3037327;
RA   Kelley M.R., Kidd S., Berg R.L., Young M.W.;
RT   "Restriction of P-element insertions at the Notch locus of Drosophila
RT   melanogaster.";
RL   Mol. Cell. Biol. 7:1545-1548(1987).
RN   [7]
RP   INTERACTION WITH DX, AND MUTANT SU42C.
RX   MEDLINE=94215489; PubMed=8162848;
RA   Diederich R.J., Matsuno K., Hing H., Artavanis-Tsakonas S.;
RT   "Cytosolic interaction between deltex and Notch ankyrin repeats
RT   implicates deltex in the Notch signaling pathway.";
RL   Development 120:473-481(1994).
RN   [8]
RP   INTERACTION WITH DX.
RX   MEDLINE=95401878; PubMed=7671825;
RA   Matsuno K., Diederich R.J., Go M.J., Blaumueller C.M.,
RA   Artavanis-Tsakonas S.;
RT   "Deltex acts as a positive regulator of Notch signaling through
RT   interactions with the Notch ankyrin repeats.";
RL   Development 121:2633-2644(1995).
RN   [9]
RP   S3 CLEAVAGE BY PSN.
RX   MEDLINE=99221487; PubMed=10206646;
RA   Struhl G., Greenwald I.;
RT   "Presenilin is required for activity and nuclear access of Notch in
RT   Drosophila.";
RL   Nature 398:522-525(1999).
RN   [10]
RP   S3 CLEAVAGE BY PSN.
RX   MEDLINE=99221488; PubMed=10206647;
RA   Ye Y., Lukinova N., Fortini M.E.;
RT   "Neurogenic phenotypes and altered Notch processing in Drosophila
RT   Presenilin mutants.";
RL   Nature 398:525-529(1999).
RN   [11]
RP   S2 CLEAVAGE BY KUZ.
RX   MEDLINE=21657146; PubMed=11799064;
```

```
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leukocyte;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=21990337; PubMed=11994479;
RA   McGreal E.P., Ikewaki N., Akatsu H., Morgan B.P., Gasque P.;
RT   "Human C1qRp is identical with CD93 and the mNI-11 antigen but does
RT   not bind C1q.";
RL   J. Immunol. 168:5222-5232(2002).
RN   [6]
RP   O-GLYCOSYLATION.
RX   MEDLINE=99192777; PubMed=10092817;
RA   Nepomuceno R.R., Ruiz S., Park M., Tenner A.J.;
RT   "C1qRP is a heavily O-glycosylated cell surface protein involved in
RT   the regulation of phagocytic activity.";
RL   J. Immunol. 162:3583-3589(1999).
CC   -!- FUNCTION: Receptor (or element of a larger receptor complex) for
CC       C1q, mannose-binding lectin (MBL2) and pulmonary surfactant
CC       protein A (SPA). May mediate the enhancement of phagocytosis in
CC       monocytes and macrophages upon interaction with soluble defense
CC       collagens. May play a role in intercellular adhesion.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: Highly expressed in endothelial cells,
CC       platelets, cells of myeloid origin, such as monocytes and
CC       neutrophils. Not expressed in cells of lymphoid origin.
CC   -!- PTM: N- and O-glycosylated.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 5 EGF-like domains.
CC   -!- CAUTION: Has been sometimes referred to as a collectin receptor.
CC   -!- CAUTION: According to Ref.5, C1q is not a ligand for C1QR1.
CC   -!- DATABASE: NAME=PROW; NOTE=PROW 3:1-6(2001);
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/guide/467246456_g.htm".
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U94333; AAB53110.1; -.
DR   EMBL; AL118508; CAC00597.1; -.
DR   EMBL; BC028075; AAH28075.1; -.
DR   HSSP; P35555; 1EMN.
DR   Genew; HGNC:15855; C1QR1.
DR   MIM; 120577; -.
DR   GO; GO:0016021; C:integral to membrane; IC.
DR   GO; GO:0004872; F:receptor activity; NAS.
DR   GO; GO:0016337; P:cell-cell adhesion; IDA.
DR   GO; GO:0042116; P:macrophage activation; NAS.
DR   GO; GO:0006909; P:phagocytosis; NAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001304; Lectin_C.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00059; lectin_c; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; FALSE_NEG.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50026; EGF_3; 3.
DR   PROSITE; PS01187; EGF_CA; 3.
KW   Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW   EGF-like domain; Lectin; Glycoprotein; Polymorphism.
FT   SIGNAL        1     21
FT   CHAIN        22    652       COMPLEMENT COMPONENT C1Q RECEPTOR.
FT   DOMAIN       24    580       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    581    601       POTENTIAL.
FT   DOMAIN      602    652       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       32    174       C-TYPE LECTIN.
FT   DOMAIN      260    301       EGF-LIKE 1.
FT   DOMAIN      302    344       EGF-LIKE 2.
FT   DOMAIN      345    384       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      385    426       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      427    468       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      594    601       POLY-LEU.
FT   DISULFID    264    275       BY SIMILARITY.
FT   DISULFID    271    285       BY SIMILARITY.
FT   DISULFID    287    300       BY SIMILARITY.
FT   DISULFID    306    317       BY SIMILARITY.
FT   DISULFID    311    328       BY SIMILARITY.
FT   DISULFID    330    343       BY SIMILARITY.
FT   DISULFID    349    358       BY SIMILARITY.
FT   DISULFID    354    367       BY SIMILARITY.
FT   DISULFID    369    383       BY SIMILARITY.
FT   DISULFID    389    400       BY SIMILARITY.
FT   DISULFID    396    409       BY SIMILARITY.
FT   DISULFID    411    425       BY SIMILARITY.
FT   DISULFID    431    443       BY SIMILARITY.
FT   DISULFID    439    452       BY SIMILARITY.
FT   DISULFID    454    467       BY SIMILARITY.
FT   CARBOHYD    325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     318    318       V -> A.
FT                                /FTId=VAR_013573.
FT   CONFLICT     22     22       T -> V (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     36     36       C -> T (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     38     39       TA -> RI (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    155    155       S -> N (IN REF. 1).
FT   CONFLICT    186    186       G -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    492    492       S -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    496    496       R -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    504    504       R -> G (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    541    541       P -> S (IN REF. 1).
SQ   SEQUENCE   652 AA;  68560 MW;  EECA0FEAC55FCAC2 CRC64;

  Query Match             11.6%;  Score 175;  DB 1;  Length 652;
  Best Local Similarity   33.1%;  Pred. No. 2.3e-06;
  Matches   60;  Conservative   10;  Mismatches   57;  Indels   54;  Gaps   13;

Qy         25 RPGRRV------CAVRAHGDPVSESFVQRVYQPFLTTCDGHRACSTYRTIYRTAYRRSPG 78
              ||| |:      || |   :| | |            | |   |
Db        288 RPGFRLLDDLVTCASR---NPCSSS-----------PCRGGATCV--------------- 318

Qy         79 LAPARPRYAC-CP-GWKRTSGLPGACGAAICQ-PPCRNGGSCVQ-PG--RCRCPAGWR-- 130
              | |    | | || |::  |          ||  ||     ||  ||  || |  |:
Db        319 LGPHGKNYTCRCPQGYQLDSSQLDCVDVDECQDSPCAQ--ECVNTPGGFRCECWVGYEPG 376
```

```
DR   EMBL; AF058897; AAC14346.1; JOINED.
DR   EMBL; AF058898; AAC14346.1; JOINED.
DR   EMBL; AF058899; AAC14346.1; JOINED.
DR   EMBL; AC004257; AAC04897.1; -.
DR   EMBL; AC004663; AAC15789.1; ALT_INIT.
DR   PIR; S78549; S78549.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:7883; NOTCH3.
DR   MIM; 600276; -.
DR   MIM; 125310; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 3.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 19.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 25.
DR   PROSITE; PS50026; EGF_3; 34.
DR   PROSITE; PS01187; EGF_CA; 16.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW   Disease mutation.
FT   SIGNAL        1     39       POTENTIAL.
FT   CHAIN        40   2321       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
FT   CHAIN      1629   2321       NOTCH EXTRACELLULAR TRUNCATION (BY
FT                                SIMILARITY).
FT   CHAIN      1662   2321       NOTCH INTRACELLULAR DOMAIN (BY
FT                                SIMILARITY).
FT   DOMAIN       40   1643       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1644   1664       POTENTIAL.
FT   DOMAIN     1665   2321       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       40     77       EGF-LIKE 1.
FT   DOMAIN       78    118       EGF-LIKE 2.
FT   DOMAIN      119    156       EGF-LIKE 3.
FT   DOMAIN      158    195       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      197    234       EGF-LIKE 5.
FT   DOMAIN      236    272       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      274    312       EGF-LIKE 7.
FT   DOMAIN      314    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      351    389       EGF-LIKE 9.
FT   DOMAIN      391    429       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      431    467       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      469    505       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      507    543       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      545    580       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      582    618       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      620    655       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      657    693       EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      695    730       EGF-LIKE 18.
FT   DOMAIN      734    770       EGF-LIKE 19.
FT   DOMAIN      771    808       EGF-LIKE 20.
FT   DOMAIN      810    847       EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      849    885       EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      887    922       EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
```

```
  Query Match             11.7%;  Score 175.5;  DB 1;  Length 2321;
  Best Local Similarity   34.5%;  Pred. No. 8.1e-06;
  Matches   50;  Conservative   14;  Mismatches   44;  Indels   37;  Gaps   11;

Qy         56 CDGHRACSTYRTIYRTAYRRSPGLAPARPRYAC-CPGWKR--TSGLPGACGAAICQPPCR 112
              | |   | :              :     |::| ||   |     ||  |       ||
Db         87 CAGRGVCQS-------------SVVAGTARFSCRCPRGFRGPDCSLPDPC----LSSPCA 129

Qy        113 NGGSC-VQP-GR--CRCPAGWRGDTCQSDVDEC----SARRGGCPQRCINTAGSYWCQCW 164
              :|  | | | ||  | || |::| :|:||||||      | ||    |:|| ||: |||
Db        130 HGARCSVGPDGRFLCSCPPGYQGRSCRSDVDECRVGEPCRHGG---TCLNTPGSFRCQCP 186

Qy        165 EGHSLSADGTLCVPKGGPPRVAPNP 189
               |::    | ||     |   ||:|
Db        187 AGYT----GPLCENPAVP--CAPSP 205


RESULT 12
CD93_HUMAN
ID   CD93_HUMAN     STANDARD;      PRT;   652 AA.
AC   Q9NPY3; O00274;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Complement component C1q receptor precursor (Complement component 1, q
DE   subcomponent, receptor 1) (C1qRp) (C1qR(p)) (C1q/MBL/SPA receptor)
DE   (CD93 antigen) (CDw93).
GN   C1QR1 OR CD93.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=97199258; PubMed=9047234;
RA   Nepomuceno R.R., Henschen-Edman A.H., Burgess W.H., Tenner A.J.;
RT   "cDNA cloning and primary structure analysis of C1qR(P), the human
RT   C1q/MBL/SPA receptor that mediates enhanced phagocytosis in vitro.";
RL   Immunity 6:119-129(1997).
RN   [2]
RP   SEQUENCE FROM N.A., AND VARIANT ALA-318.
RX   MEDLINE=21640567; PubMed=11781389;
RA   Steinberger P., Szekeres A., Wille S., Stockl J., Selenko N.,
RA   Prager E., Staffler G., Madic O., Stockinger H., Knapp W.;
RT   "Identification of human CD93 as the phagocytic C1q receptor (C1qRp)
RT   by expression cloning.";
RL   J. Leukoc. Biol. 71:133-140(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052;
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
```

```
                    |  |:  |      ||  :|:  | ||| || | || |  |: |  |:
Db         131 DDHCLCQKGYIGTH-----CGQPVCESGCLNGGRCVAPNRCACTYGFTGPQCERDYRTGP 185

Qy         136 -----------------------------------------------------------S 136

Db         186 CFTVISNQMCQGQLSGIVCTKQLCCATVGRAWGHPCEMCPAQPHPCRRGFIPNIRTGACQ 245

Qy         137 DVDECSARRGGCP-QRCINTAGSYWCQCWEGHSLSADGTLC-----------VPKGG--- 181
               ||||| |  | |    |||| ||: |:|  || |:     |           : :||
Db         246 DVDECQAIPGLCQGGNCINTVGSFECKCPAGHKLNEVSQKCEDIDECSTIPGICEGGECT 305

Qy         182 ----------PPRVAPNPTG 191
                         ||    :| |
Db         306 NTVSSYFCKCPPGFYTSPDG 325


RESULT 11
NTC3_HUMAN
ID   NTC3_HUMAN     STANDARD;      PRT;  2321 AA.
AC   Q9UM47; Q9UEB3; Q9UPL3; Q9Y6L8;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 3 precursor (Notch 3).
GN   NOTCH3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97032728; PubMed=8878478;
RA   Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
RA   Alamowitch S., Domenga V., Cecillion M., Marechal E., Maciazek J.,
RA   Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissenbach J.,
RA   Bach J.-F., Bousser M.-G., Tournier-Lasserve E.;
RT   "Notch3 mutations in CADASIL, a hereditary adult-onset condition
RT   causing stroke and dementia.";
RL   Nature 383:707-710(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Gunel M., Artavanis-Tsakonas S.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA   Burkhart-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA   Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,
RA   Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA   Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA   Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA   Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA   Carrano A.V.;
RT   "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
RT   19p13.1.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   VARIANTS CADASIL TYR-49; CYS-71; CYS-90; CYS-110; CYS-133; CYS-141;
RP   ARG-146; CYS-153; CYS-169; CYS-171; CYS-182; ARG-185; SER-212;
RP   GLY-222; TYR-224; CYS-258; TYR-542; CYS-558; CYS-578; CYS-728;
RP   CYS-985; CYS-1006; CYS-1031; CYS-1231 AND ARG-1261, AND VARIANTS
RP   ARG-170; LEU-496; GLN-1133; MET-1183 AND ALA-2223.
RX   MEDLINE=98049753; PubMed=9388399;
RA   Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA   Vayssiere C., Cruaud C., Maciazek J., Weissenbach J., Bousser M.-G.,
RA   Bach J.-F., Tournier-Lasserve E.;
RT   "Strong clustering and stereotyped nature of Notch3 mutations in
RT   CADASIL patients.";
RL   Lancet 350:1511-1515(1997).
RN   [5]
RP   VARIANT CADASIL 114-GLY--PRO-120 DEL.
RX   MEDLINE=20264473; PubMed=10802807;
RA   Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
RA   Ruchoux M.M., Lucas C., Leys D., Bousser M.-G., Tournier-Lasserve E.;
RT   "Splice site mutation causing a seven amino acid Notch3 in-frame
RT   deletion in CADASIL.";
RL   Neurology 54:1874-1875(2000).
RN   [6]
RP   IDENTIFICATION OF LIGANDS.
RX   MEDLINE=99180765; PubMed=10079256;
RA   Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA   Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT   "Human ligands of the Notch receptor.";
RL   Am. J. Pathol. 154:785-794(1999).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
CC       tissues.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- DISEASE: Defects in NOTCH3 are associated with cerebral autosomal
CC       dominant arteriopathy with subcortical infarcts and
CC       leukoencephalopathy (CADASIL) [MIM:125310]. CADASIL causes a type
CC       of stroke and dementia of which key features include recurrent
CC       subcortical ischemic events and vascular dementia.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 34 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U97669; AAB91371.1; -.
DR   EMBL; AF058900; AAC14346.1; -.
DR   EMBL; AF058881; AAC14346.1; JOINED.
DR   EMBL; AF058882; AAC14346.1; JOINED.
DR   EMBL; AF058883; AAC14346.1; JOINED.
DR   EMBL; AF058884; AAC14346.1; JOINED.
DR   EMBL; AF058885; AAC14346.1; JOINED.
DR   EMBL; AF058886; AAC14346.1; JOINED.
DR   EMBL; AF058887; AAC14346.1; JOINED.
DR   EMBL; AF058888; AAC14346.1; JOINED.
DR   EMBL; AF058889; AAC14346.1; JOINED.
DR   EMBL; AF058890; AAC14346.1; JOINED.
DR   EMBL; AF058891; AAC14346.1; JOINED.
DR   EMBL; AF058892; AAC14346.1; JOINED.
DR   EMBL; AF058893; AAC14346.1; JOINED.
DR   EMBL; AF058894; AAC14346.1; JOINED.
DR   EMBL; AF058895; AAC14346.1; JOINED.
DR   EMBL; AF058896; AAC14346.1; JOINED.
```

```
RN   [8]
RP   STRUCTURE BY NMR OF 2124-2205.
RX   MEDLINE=96222301; PubMed=8653794;
RA   Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,
RA   Handford P.A.;
RT   "Solution structure of a pair of calcium-binding epidermal growth
RT   factor-like domains: implications for the Marfan syndrome and other
RT   genetic disorders.";
RL   Cell 85:597-605(1996).
RN   [9]
RP   REVIEW ON MFS VARIANTS.
RX   MEDLINE=96174615; PubMed=8594563;
RA   Collod G., Beroud C., Soussi T., Junien C., Boileau C.;
RT   "Software and database for the analysis of mutations in the human
RT   FBN1 gene.";
RL   Nucleic Acids Res. 24:137-141(1996).
RN   [10]
RP   REVIEW ON MFS VARIANTS.
RX   MEDLINE=97169383; PubMed=9016526;
RA   Collod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,
RA   Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,
RA   Richards R.I., Wang W., Junien C., Boileau C.;
RT   "Marfan Database (second edition): software and database for the
RT   analysis of mutations in the human FBN1 gene.";
RL   Nucleic Acids Res. 25:147-150(1997).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=98062175; PubMed=9401003;
RA   Hayward C., Brock D.J.H.;
RT   "Fibrillin-1 mutations in Marfan syndrome and other type-1
RT   fibrillinopathies.";
RL   Hum. Mutat. 10:415-423(1997).
RN   [12]
RP   VARIANT MFS PRO-1137.
RX   MEDLINE=91304569; PubMed=1852208;
RA   Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
RA   Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
RA   Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;
RT   "Marfan syndrome caused by a recurrent de novo missense mutation in
RT   the fibrillin gene.";
RL   Nature 352:337-339(1991).
RN   [13]
RP   VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.
RX   MEDLINE=93250834; PubMed=1301946;
RA   Dietz H.C., Saraiva J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
RT   "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
RT   patients at cysteine residues in EGF-like domains.";
RL   Hum. Mutat. 1:366-374(1992).
RN   [14]
RP   VARIANT MFS SER-2307.
RX   MEDLINE=92235290; PubMed=1569206;
RA   Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,
RA   Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
RT   "Marfan phenotype variability in a family segregating a missense
RT   mutation in the epidermal growth factor-like motif of the fibrillin
RT   gene.";
RL   J. Clin. Invest. 89:1674-1680(1992).
RN   [15]
RP   VARIANTS MFS ILE-548 AND ALA-723.
RX   MEDLINE=94010946; PubMed=8406497;
RA   Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
RA   Pyeritz R.E., Francomano C.A.;
RT   "Four novel FBN1 mutations: significance for mutant transcript level
RT   and EGF-like domain calcium binding in the pathogenesis of Marfan
RT   syndrome.";
RL   Genomics 17:468-475(1993).
RN   [16]
RP   VARIANT MFS SER-2144.
RX   MEDLINE=93278402; PubMed=8504310;
RA   Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
RT   "A novel fibrillin mutation in the Marfan syndrome which could
RT   disrupt calcium binding of the epidermal growth factor-like module.";
RL   Hum. Mol. Genet. 2:475-477(1993).
```

```
RN   [17]
RP   VARIANTS MFS ARG-862; TYR-1117; PRO-1137 AND PHE-1589, AND VARIANT
RP   ALA-1148.
RX   MEDLINE=94108431; PubMed=8281141;
RA   Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
RA   Berg M.A., Miller D.C., Francke U.;
RT   "Mutation screening of complete fibrillin-1 coding sequence: report
RT   of five new mutations, including two in 8-cysteine domains.";
RL   Hum. Mol. Genet. 2:1813-1821(1993).
RN   [18]
RP   VARIANTS MFS GLY-217 AND ARG-2627.
RX   MEDLINE=95067970; PubMed=7977366;
RA   Karttunen L., Raghunath M., Loennqvist L., Peltonen L.;
RT   "A compound-heterozygous Marfan patient: two defective fibrillin
RT   alleles result in a lethal phenotype.";
RL   Am. J. Hum. Genet. 55:1083-1091(1994).
RN   [19]
RP   VARIANT EL LYS-2447.
RX   MEDLINE=94245249; PubMed=8188302;
RA   Lonnqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
RA   Peltonen L.;
RT   "A novel mutation of the fibrillin gene causing ectopia lentis.";
RL   Genomics 19:573-576(1994).
RN   [20]
RP   VARIANT MFS CYS-627.
RX   MEDLINE=94272487; PubMed=8004112;
RA   Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
RT   "Two novel mutations and a neutral polymorphism in EGF-like domains
RT   of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
RT   syndrome patients.";
RL   Hum. Mol. Genet. 3:373-375(1994).
RN   [21]
RP   VARIANT MFS CYS-122.
RX   MEDLINE=94314977; PubMed=8040326;
RA   Stahl-Hallengren C., Ukkonen T., Kainulainen K., Kristofersson U.,
RA   Saxne T., Tornqvist K., Peltonen L.;
RT   "An extra cysteine in one of the non-calcium-binding epidermal growth
RT   factor-like motifs of the FBN1 polypeptide is connected to a novel
RT   variant of Marfan syndrome.";
RL   J. Clin. Invest. 94:709-713(1994).
RN   [22]
RP   VARIANT MFS TYR-1223.
RX   MEDLINE=94351682; PubMed=8071963;
RA   Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
RT   "A new missense mutation of fibrillin in a patient with Marfan
RT   syndrome.";
RL   J. Med. Genet. 31:338-339(1994).
RN   [23]
RP   VARIANT MFS HIS-1170.
RX   MEDLINE=95174777; PubMed=7870075;
RA   Hayward C., Porteous M.E.M., Brock D.J.H.;
RT   "A novel mutation in the fibrillin gene (FBN1) in familial
RT   arachnodactyly.";
RL   Mol. Cell. Probes 8:325-327(1994).
RN   [24]
RP   VARIANTS MFS GLY-217; ASN-1023; ARG-1074; TYR-1242; ARG-1513;
RP   GLU-2127; TRP-2151; LYS-2447 AND ARG-2511.
RX   MEDLINE=94184368; PubMed=8136837;
RA   Kainulainen K., Karttunen L., Puhakka L., Sakai L., Peltonen L.;
RT   "Mutations in the fibrillin gene responsible for dominant ectopia
RT   lentis and neonatal Marfan syndrome.";
RL   Nat. Genet. 6:64-69(1994).
RN   [25]

  Query Match             11.7%;  Score 176.5;  DB 1;  Length 2871;
  Best Local Similarity   22.7%;  Pred. No. 8.5e-06;
  Matches   59;  Conservative   19;  Mismatches   79;  Indels  103;  Gaps    8;

Qy          26 PGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR---ACSTYRTIYRTAYRRSPGLAPA 82
               ||   | |        : |  |   | : ||   :   :| : |:|     |   | : :
Db          75 PGGNQCIVPICRHSCGDGFCSR---PNMCTCPSGQIAPSCGS-RSIQHCNIRCMNGGSCS 130

Qy          83 RPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQ------- 135
```

```
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D86747; BAA13167.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 3.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    816       NEL PROTEIN.
FT   DOMAIN       30    258       TSP N-TERMINAL.
FT   DOMAIN      272    331       VWFC 1.
FT   DOMAIN      332    396       VWFC 2.
FT   DOMAIN      397    439       EGF-LIKE 1.
FT   DOMAIN      440    481       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      482    522       EGF-LIKE 3.
FT   DOMAIN      521    553       EGF-LIKE 4.
FT   DOMAIN      555    601       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      602    637       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      638    693       VWFC 3.
FT   DOMAIN      698    756       VWFC 4.
FT   DOMAIN      758    813       VWFC 5.
FT   DISULFID    401    413       BY SIMILARITY.
FT   DISULFID    407    422       BY SIMILARITY.
FT   DISULFID    424    438       BY SIMILARITY.
FT   DISULFID    444    457       BY SIMILARITY.
FT   DISULFID    451    466       BY SIMILARITY.
FT   DISULFID    468    480       BY SIMILARITY.
FT   DISULFID    486    499       BY SIMILARITY.
FT   DISULFID    493    508       BY SIMILARITY.
FT   DISULFID    510    521       BY SIMILARITY.
FT   DISULFID    525    535       BY SIMILARITY.
FT   DISULFID    529    541       BY SIMILARITY.
FT   DISULFID    543    552       BY SIMILARITY.
FT   DISULFID    559    572       BY SIMILARITY.
FT   DISULFID    566    581       BY SIMILARITY.
FT   DISULFID    583    600       BY SIMILARITY.
FT   DISULFID    606    619       BY SIMILARITY.
FT   DISULFID    613    628       BY SIMILARITY.
FT   DISULFID    630    636       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    635    635       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   816 AA;  90968 MW;  BFCDBC983C02F831 CRC64;

  Query Match             11.7%;  Score 176.5;  DB 1;  Length 816;
  Best Local Similarity   39.8%;  Pred. No. 2.2e-06;
  Matches   33;  Conservative   12;  Mismatches   33;  Indels    5;  Gaps    2;

Qy         100 GACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQR--CINTAG 157
               |    | |:  |||||:|:    | || |: | :|::|:||||     |  |  |||  |
Db         518 GTVCKAFCKDGCRNGGACIASNVCACPQGFTGPSCETDIDECSDGFVQCDSRANCINLPG 577

Qy         158 SYWCQCWEGHSLSADGTLCVPKG 180
                | |:| :|:    |  :  | |
Db         578 WYHCECRDGYH---DNGMFSPSG 597


RESULT 10
FBN1_HUMAN
ID   FBN1_HUMAN     STANDARD;      PRT;  2871 AA.
AC   P35555;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Fibrillin 1 precursor.
GN   FBN1 OR FBN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=93372860; PubMed=8364578;
RA   Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA   Pangilinan T., Bonadio J.;
RT   "Genomic organization of the sequence coding for fibrillin, the
RT   defective gene product in Marfan syndrome.";
RL   Hum. Mol. Genet. 2:961-968(1993).
RN   [2]
RP   SEQUENCE OF 1-932 FROM N.A.
RC   TISSUE=Fibroblast, and Placenta;
RX   MEDLINE=94010947; PubMed=7691719;
RA   Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT   "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT   multidomain structure and alternatively spliced exons at the 5'
RT   end.";
RL   Genomics 17:476-484(1993).
RN   [3]
RP   SEQUENCE OF 899-2871 FROM N.A.
RX   MEDLINE=91304568; PubMed=1852207;
RA   Maslen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT   "Partial sequence of a candidate gene for the Marfan syndrome.";
RL   Nature 352:334-337(1991).
RN   [4]
RP   SEQUENCE OF 813-1313 FROM N.A.
RX   MEDLINE=91304567; PubMed=1852206;
RA   Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA   Tsipouras P., Ramirez F., Hollister D.W.;
RT   "Linkage of Marfan syndrome and a phenotypically related disorder to
RT   two different fibrillin genes.";
RL   Nature 352:330-334(1991).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=91317849; PubMed=1860873;
RA   Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;
RT   "Purification and partial characterization of fibrillin, a cysteine-
RT   rich structural component of connective tissue microfibrils.";
RL   J. Biol. Chem. 266:14763-14770(1991).
RN   [6]
RP   STRUCTURE BY NMR OF 2054-2125.
RX   MEDLINE=98031893; PubMed=9362480;
RA   Yuan X., Downing A.K., Knott V., Handford P.A.;
RT   "Solution structure of the transforming growth factor beta-binding
RT   protein-like module, a domain associated with matrix fibrils.";
RL   EMBO J. 16:6659-6666(1997).
RN   [7]
RP   STRUCTURE BY NMR OF 2124-2205.
RX   MEDLINE=96144829; PubMed=8568869;
RA   Knott V., Downing A.K., Cardy C.M., Handford P.A.;
RT   "Calcium binding properties of an epidermal growth factor-like domain
RT   pair from human fibrillin-1.";
RL   J. Mol. Biol. 255:22-27(1996).
```

```
RA   Abe T., Matsuhashi S., Ting K.;
RT   "Biochemical characterization and expression analysis of neural
RT   thrombospondin-1-like proteins NELL1 and NELL2.";
RL   Biochem. Biophys. Res. Commun. 265:79-86(1999).
CC   -!- SUBUNIT: Homotrimer. Binds to PKC beta-1.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -!- SIMILARITY: Contains 5 VWFC domains.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U48246; AAC72252.1; -.
DR   PIR; T10756; T10756.
DR   HSSP; P35555; 1EMN.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50026; EGF_3; 5.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 2.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    810       PROTEIN KINASE C-BINDING PROTEIN NELL1.
FT   DOMAIN       81    230       TSP N-TERMINAL.
FT   DOMAIN      271    332       VWFC 1.
FT   DOMAIN      335    390       VWFC 2.
FT   DOMAIN      391    433       EGF-LIKE 1.
FT   DOMAIN      434    475       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      476    516       EGF-LIKE 3.
FT   DOMAIN      515    547       EGF-LIKE 4.
FT   DOMAIN      549    595       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      596    631       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      632    687       VWFC 3.
FT   DOMAIN      692    750       VWFC 4.
FT   DOMAIN      752    807       VWFC 5.
FT   DISULFID    395    407       BY SIMILARITY.
FT   DISULFID    401    416       BY SIMILARITY.
FT   DISULFID    418    432       BY SIMILARITY.
FT   DISULFID    438    451       BY SIMILARITY.
FT   DISULFID    445    460       BY SIMILARITY.
FT   DISULFID    462    474       BY SIMILARITY.
FT   DISULFID    480    493       BY SIMILARITY.
FT   DISULFID    487    502       BY SIMILARITY.
FT   DISULFID    504    515       BY SIMILARITY.
FT   DISULFID    519    529       BY SIMILARITY.
FT   DISULFID    523    535       BY SIMILARITY.
FT   DISULFID    537    546       BY SIMILARITY.
FT   DISULFID    553    566       BY SIMILARITY.
FT   DISULFID    560    575       BY SIMILARITY.
FT   DISULFID    577    594       BY SIMILARITY.
FT   DISULFID    600    613       BY SIMILARITY.
FT   DISULFID    607    622       BY SIMILARITY.
FT   DISULFID    624    630       BY SIMILARITY.
FT   CARBOHYD     40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    224    224       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    294    294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    511    511       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    562    562       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    609    609       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    708    708       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   810 AA;  89212 MW;  46F09C466AF9AB0B CRC64;

  Query Match             12.0%;  Score 180;  DB 1;  Length 810;
  Best Local Similarity   36.5%;  Pred. No. 1.2e-06;
  Matches   35;  Conservative   13;  Mismatches   38;  Indels   10;  Gaps    3;

Qy         88 CCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGG 147
              | ||:       |    | |:  || ||:|| | :| ||:|: |  |: |:|||:
Db        504 CQPGYVGN----GTICKAFCEEGCRYGGTCVAPNKCVCPSGFTGSHCEKDIDECAEGFVE 559

Qy        148 CP--QRCINTAGSYWCQCWEGH----SLSADGTLCV 177
              |    ||:|  | | |:|  |     : |  |  |:
Db        560 CHNYSRCVNLPGWYHCECRSGFHDDGTYSLSGESCI 595


RESULT 9
NEL_CHICK
ID   NEL_CHICK      STANDARD;      PRT;   816 AA.
AC   Q90827;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   NEL protein precursor (93 kDa protein).
GN   NEL.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=95383734; PubMed=7655083;
RA   Matsuhashi S., Noji S., Koyama E., Myokai F., Ohuchi H.,
RA   Taniguchi S., Hori K.;
RT   "New gene, nel, encoding a M(r) 93 K protein with EGF-like repeats is
RT   strongly expressed in neural tissues of early stage chick embryos.";
RL   Dev. Dyn. 203:212-222(1995).
RN   [2]
RP   TSP N-TERMINAL DOMAIN.
RX   MEDLINE=98153258; PubMed=9480764;
RA   Beckmann G., Hanke J., Bork P., Reich J.;
RT   "Merging extracellular domains: fold prediction for laminin G-like
RT   and amino-terminal thrombospondin-like modules based on homology to
RT   pentraxins.";
RL   J. Mol. Biol. 275:725-730(1998).
CC   -!- TISSUE SPECIFICITY: Strongly expressed in early embryonic neural
CC       tissues (brain, spinal cord, dorsal root ganglia); less in other
CC       tissues such as cells around cartilage, myocardium, lung
CC       mesenchymal cells, and liver. After hatching expression is
CC       restricted to neural tissues including retina.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -!- SIMILARITY: Contains 5 VWFC domains.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
Qy          148 C--PQRCINTAGSYWCQCWEGH----SLSADGTLCV 177
                |    ||:|  | | |:|  |     : |  |  |:
Db          560 CHNHSRCVNLPGWYHCECRSGFHDDGTYSLSGESCI 595

RESULT 7
NEL2_RAT
ID   NEL2_RAT       STANDARD;      PRT;   816 AA.
AC   Q62918;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein kinase C-binding protein NELL2 precursor (NEL-like protein 2).
GN   NELL2 OR NEL.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=20017976; PubMed=10548494;
RA   Kuroda S., Oyasu M., Kawakami M., Kanayama N., Tanizawa K., Saito N.,
RA   Abe T., Matsuhashi S., Ting K.;
RT   "Biochemical characterization and expression analysis of neural
RT   thrombospondin-1-like proteins NELL1 and NELL2.";
RL   Biochem. Biophys. Res. Commun. 265:79-86(1999).
RN   [2]
RP   TSP N-TERMINAL DOMAIN.
RX   MEDLINE=98153258; PubMed=9480764;
RA   Beckmann G., Hanke J., Bork P., Reich J.;
RT   "Merging extracellular domains: fold prediction for laminin G-like
RT   and amino-terminal thrombospondin-like modules based on homology to
RT   pentraxins.";
RL   J. Mol. Biol. 275:725-730(1998).
CC   -!- SUBUNIT: Homotrimer. Binds to PKC beta-1.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -!- SIMILARITY: Contains 5 VWFC domains.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U48245; AAC72245.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 3.
DR   SMART; SM00179; EGF_CA; 3.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 3.
```

```
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    816       PROTEIN KINASE C-BINDING PROTEIN NELL2.
FT   DOMAIN       30    258       TSP N-TERMINAL.
FT   DOMAIN      272    331       VWFC 1.
FT   DOMAIN      332    396       VWFC 2.
FT   DOMAIN      397    439       EGF-LIKE 1.
FT   DOMAIN      440    481       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      482    522       EGF-LIKE 3.
FT   DOMAIN      521    553       EGF-LIKE 4.
FT   DOMAIN      555    601       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      602    637       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      638    693       VWFC 3.
FT   DOMAIN      698    756       VWFC 4.
FT   DOMAIN      758    813       VWFC 5.
FT   DISULFID    401    413       BY SIMILARITY.
FT   DISULFID    407    422       BY SIMILARITY.
FT   DISULFID    424    438       BY SIMILARITY.
FT   DISULFID    444    457       BY SIMILARITY.
FT   DISULFID    451    466       BY SIMILARITY.
FT   DISULFID    468    480       BY SIMILARITY.
FT   DISULFID    486    499       BY SIMILARITY.
FT   DISULFID    493    508       BY SIMILARITY.
FT   DISULFID    510    521       BY SIMILARITY.
FT   DISULFID    525    535       BY SIMILARITY.
FT   DISULFID    529    541       BY SIMILARITY.
FT   DISULFID    543    552       BY SIMILARITY.
FT   DISULFID    559    572       BY SIMILARITY.
FT   DISULFID    566    581       BY SIMILARITY.
FT   DISULFID    583    600       BY SIMILARITY.
FT   DISULFID    606    619       BY SIMILARITY.
FT   DISULFID    613    628       BY SIMILARITY.
FT   DISULFID    630    636       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    635    635       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   816 AA;  90952 MW;  A999F76078060D6B CRC64;

  Query Match             12.0%;  Score 180.5;  DB 1;  Length 816;
  Best Local Similarity   36.8%;  Pred. No. 1.1e-06;
  Matches   35;  Conservative   14;  Mismatches   37;  Indels    9;  Gaps    3;

Qy           88 CCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGG 147
                | ||:       |    | |:  |:|||:|:    | || |: | :|::|:||||
Db          510 CKPGYTGN----GTTCKAFCKDGCKNGGACIAANVCACPQGFTGPSCETDIDECSEGFVQ 565

Qy          148 CPQR--CINTAGSYWCQCWEGHSLSADGTLCVPKG 180
                |  |  |||  | | |:| :|:    |  :  | |
Db          566 CDSRANCINLPGWYHCECRDGYH---DNGMFAPGG 597


RESULT 8
NEL1_RAT
ID   NEL1_RAT       STANDARD;      PRT;   810 AA.
AC   Q62919;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein kinase C-binding protein NELL1 precursor (NEL-like protein 1).
GN   NELL1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=20017976; PubMed=10548494;
RA   Kuroda S., Oyasu M., Kawakami M., Kanayama N., Tanizawa K., Saito N.,
```

```
FT   CARBOHYD    635    635       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   816 AA;  91346 MW;  89370B987DC7A324 CRC64;

  Query Match             12.1%;  Score 182.5;  DB 1;  Length 816;
  Best Local Similarity   37.9%;  Pred. No. 7.5e-07;
  Matches   36;  Conservative   13;  Mismatches   37;  Indels    9;  Gaps    3;

Qy          88 CCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGG 147
               | ||:       |    | |:  |||||:|:    | || |: | :|::|:||||
Db         510 CKPGYTGN----GTTCKAFCKDGCRNGGACIAANVCACPQGFTGPSCETDIDECSDGFVQ 565

Qy         148 CPQR--CINTAGSYWCQCWEGHSLSADGTLCVPKG 180
               |  |  |||  | | |:| :|:    |  :  | |
Db         566 CDSRANCINLPGWYHCECRDGYH---DNGMFSPSG 597


RESULT 6
NEL1_HUMAN
ID   NEL1_HUMAN     STANDARD;      PRT;   810 AA.
AC   Q92832; Q9Y472;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein kinase C-binding protein NELL1 precursor (NEL-like protein 1)
DE   (Nel-related protein 1).
GN   NELL1 OR NRP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97131504; PubMed=8975702;
RA   Watanabe T.K., Katagiri T., Suzuki M., Shimizu F., Fujiwara T.,
RA   Kanemoto N., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.;
RT   "Cloning and characterization of two novel human cDNAs (NELL1 and
RT   NELL2) encoding proteins with six EGF-like repeats.";
RL   Genomics 38:273-276(1996).
RN   [2]
RP   SEQUENCE OF 383-810 FROM N.A.
RA   Ting K., Vastardis H., Mulliken J.B., Bertolami C., Wen Z.,
RA   Young M., Tieu A., Kwong E.;
RT   "Nel homolog gene expression in craniofacial anomalies.";
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: Homotrimer. Binds to PKC beta-1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- DISEASE: Expressed in craniofacial anomalies.
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -!- SIMILARITY: Contains 5 VWFC domains.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC       in positions 427 and 771.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D83017; BAA11680.1; -.
DR   EMBL; U57523; AAB06946.1; ALT_FRAME.
DR   HSSP; P07204; 1ADX.
DR   Genew; HGNC:7750; NELL1.
DR   MIM; 602319; -.
DR   GO; GO:0007399; P:neurogenesis; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
```

```
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50026; EGF_3; 5.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 2.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17    810       PROTEIN KINASE C-BINDING PROTEIN NELL1.
FT   DOMAIN       81    230       TSP N-TERMINAL.
FT   DOMAIN      271    332       VWFC 1.
FT   DOMAIN      335    390       VWFC 2.
FT   DOMAIN      391    433       EGF-LIKE 1.
FT   DOMAIN      434    475       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      476    516       EGF-LIKE 3.
FT   DOMAIN      515    547       EGF-LIKE 4.
FT   DOMAIN      549    595       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      596    631       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      632    687       VWFC 3.
FT   DOMAIN      692    750       VWFC 4.
FT   DOMAIN      752    807       VWFC 5.
FT   DISULFID    395    407       BY SIMILARITY.
FT   DISULFID    401    416       BY SIMILARITY.
FT   DISULFID    418    432       BY SIMILARITY.
FT   DISULFID    438    451       BY SIMILARITY.
FT   DISULFID    445    460       BY SIMILARITY.
FT   DISULFID    462    474       BY SIMILARITY.
FT   DISULFID    480    493       BY SIMILARITY.
FT   DISULFID    487    502       BY SIMILARITY.
FT   DISULFID    504    515       BY SIMILARITY.
FT   DISULFID    519    529       BY SIMILARITY.
FT   DISULFID    523    535       BY SIMILARITY.
FT   DISULFID    537    546       BY SIMILARITY.
FT   DISULFID    553    566       BY SIMILARITY.
FT   DISULFID    560    575       BY SIMILARITY.
FT   DISULFID    577    594       BY SIMILARITY.
FT   DISULFID    600    613       BY SIMILARITY.
FT   DISULFID    607    622       BY SIMILARITY.
FT   DISULFID    624    630       BY SIMILARITY.
FT   CARBOHYD     40     40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    224    224       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    294    294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    372    372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    511    511       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    562    562       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    609    609       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    708    708       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    732    732       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    758    758       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    383    383       N -> D (IN REF. 2).
FT   CONFLICT    573    573       Y -> H (IN REF. 2).
FT   CONFLICT    626    626       S -> C (IN REF. 2).
SQ   SEQUENCE   810 AA;  89606 MW;  549465EA3F7AEED0 CRC64;

  Query Match             12.1%;  Score 182;  DB 1;  Length 810;
  Best Local Similarity   37.5%;  Pred. No. 8.1e-07;
  Matches   36;  Conservative   12;  Mismatches   38;  Indels   10;  Gaps    3;

Qy          88 CCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGG 147
               | ||:       |    | |:  || ||:|| | :| ||:|: |  |: |:||||
Db         504 CKPGYVGN----GTICRAFCEEGCRYGGTCVAPNKCVCPSGFTGSHCEKDIDECSEGIIE 559
```

```
                                            ||||| |  | |    |||| ||: |:|
Db        246 AWGHPCEMCPAQPQPCRPGFIPNIRTGACQDVDECQAIPGLCQGGNCINTVGSFECRCPA 305

Qy        166 GHSLSADGTLC 176
              ||  |     |
Db        306 GHKQSETTQKC 316


RESULT 5
NEL2_HUMAN
ID   NEL2_HUMAN     STANDARD;      PRT;   816 AA.
AC   Q99435;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein kinase C-binding protein NELL2 precursor (NEL-like protein 2)
DE   (Nel-related protein 2).
GN   NELL2 OR NRP2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97131504; PubMed=8975702;
RA   Watanabe T.K., Katagiri T., Suzuki M., Shimizu F., Fujiwara T.,
RA   Kanemoto N., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.;
RT   "Cloning and characterization of two novel human cDNAs (NELL1 and
RT   NELL2) encoding proteins with six EGF-like repeats.";
RL   Genomics 38:273-276(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Duodenum;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   TSP N-TERMINAL DOMAIN.
RX   MEDLINE=98153258; PubMed=9480764;
RA   Beckmann G., Hanke J., Bork P., Reich J.;
RT   "Merging extracellular domains: fold prediction for laminin G-like
RT   and amino-terminal thrombospondin-like modules based on homology to
RT   pentraxins.";
RL   J. Mol. Biol. 275:725-730(1998).
CC   -!- SUBUNIT: Homotrimer. Binds to PKC beta-1 (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -!- SIMILARITY: Contains 5 VWFC domains.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D83018; BAA11681.1; -.
DR   EMBL; BC020544; AAH20544.1; -.
DR   HSSP; P00740; 1EDM.
DR   Genew; HGNC:7751; NELL2.
DR   MIM; 602320; -.
DR   GO; GO:0005576; C:extracellular; NAS.
DR   GO; GO:0005509; F:calcium ion binding; NAS.
DR   GO; GO:0005198; F:structural molecule activity; NAS.
DR   GO; GO:0007155; P:cell adhesion; NAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 3.
DR   SMART; SM00179; EGF_CA; 3.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 3.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    816       PROTEIN KINASE C-BINDING PROTEIN NELL2.
FT   DOMAIN       30    258       TSP N-TERMINAL.
FT   DOMAIN      272    331       VWFC 1.
FT   DOMAIN      332    396       VWFC 2.
FT   DOMAIN      397    439       EGF-LIKE 1.
FT   DOMAIN      440    481       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      482    522       EGF-LIKE 3.
FT   DOMAIN      521    553       EGF-LIKE 4.
FT   DOMAIN      555    601       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      602    637       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      638    693       VWFC 3.
FT   DOMAIN      698    756       VWFC 4.
FT   DOMAIN      758    813       VWFC 5.
FT   DISULFID    401    413       BY SIMILARITY.
FT   DISULFID    407    422       BY SIMILARITY.
FT   DISULFID    424    438       BY SIMILARITY.
FT   DISULFID    444    457       BY SIMILARITY.
FT   DISULFID    451    466       BY SIMILARITY.
FT   DISULFID    468    480       BY SIMILARITY.
FT   DISULFID    486    499       BY SIMILARITY.
FT   DISULFID    493    508       BY SIMILARITY.
FT   DISULFID    510    521       BY SIMILARITY.
FT   DISULFID    525    535       BY SIMILARITY.
FT   DISULFID    529    541       BY SIMILARITY.
FT   DISULFID    543    552       BY SIMILARITY.
FT   DISULFID    559    572       BY SIMILARITY.
FT   DISULFID    566    581       BY SIMILARITY.
FT   DISULFID    583    600       BY SIMILARITY.
FT   DISULFID    606    619       BY SIMILARITY.
FT   DISULFID    613    628       BY SIMILARITY.
FT   DISULFID    630    636       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   DOMAIN       568    609       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN       610    650       EGF-LIKE 9, CALCIUM-BINDING.
FT   DOMAIN       651    691       EGF-LIKE 10, CALCIUM-BINDING.
FT   REPEAT       692    760       TGFBP 2.
FT   DOMAIN       761    802       EGF-LIKE 11, CALCIUM-BINDING.
FT   DOMAIN       803    844       EGF-LIKE 12, CALCIUM-BINDING.
FT   DOMAIN       845    883       EGF-LIKE 13, CALCIUM-BINDING.
FT   DOMAIN       948    989       EGF-LIKE 14, CALCIUM-BINDING.
FT   REPEAT       990   1065       TGFBP 3.
FT   DOMAIN      1066   1107       EGF-LIKE 15, CALCIUM-BINDING.
FT   DOMAIN      1108   1150       EGF-LIKE 16, CALCIUM-BINDING.
FT   DOMAIN      1151   1192       EGF-LIKE 17, CALCIUM-BINDING.
FT   DOMAIN      1193   1234       EGF-LIKE 18, CALCIUM-BINDING.
FT   DOMAIN      1235   1275       EGF-LIKE 19, CALCIUM-BINDING.
FT   DOMAIN      1276   1317       EGF-LIKE 20, CALCIUM-BINDING.
FT   DOMAIN      1318   1359       EGF-LIKE 21, CALCIUM-BINDING.
FT   DOMAIN      1360   1400       EGF-LIKE 22, CALCIUM-BINDING.
FT   DOMAIN      1401   1441       EGF-LIKE 23, CALCIUM-BINDING.
FT   DOMAIN      1442   1483       EGF-LIKE 24, CALCIUM-BINDING.
FT   DOMAIN      1484   1524       EGF-LIKE 25, CALCIUM-BINDING.
FT   DOMAIN      1525   1565       EGF-LIKE 26, CALCIUM-BINDING.
FT   REPEAT      1566   1642       TGFBP 4.
FT   DOMAIN      1643   1684       EGF-LIKE 27, CALCIUM-BINDING.
FT   DOMAIN      1685   1726       EGF-LIKE 28, CALCIUM-BINDING.
FT   REPEAT      1727   1800       TGFBP 5.
FT   DOMAIN      1801   1842       EGF-LIKE 29, CALCIUM-BINDING.
FT   DOMAIN      1843   1884       EGF-LIKE 30, CALCIUM-BINDING.
FT   DOMAIN      1885   1926       EGF-LIKE 31, CALCIUM-BINDING.
FT   DOMAIN      1927   1965       EGF-LIKE 32, CALCIUM-BINDING.
FT   DOMAIN      1966   2008       EGF-LIKE 33, CALCIUM-BINDING.
FT   DOMAIN      2009   2048       EGF-LIKE 34, CALCIUM-BINDING.
FT   DOMAIN      2049   2090       EGF-LIKE 35, CALCIUM-BINDING.
FT   REPEAT      2091   2163       TGFBP 6.
FT   DOMAIN      2164   2205       EGF-LIKE 36, CALCIUM-BINDING.
FT   DOMAIN      2206   2245       EGF-LIKE 37, CALCIUM-BINDING.
FT   DOMAIN      2246   2286       EGF-LIKE 38, CALCIUM-BINDING.
FT   DOMAIN      2287   2330       EGF-LIKE 39, CALCIUM-BINDING.
FT   DOMAIN      2331   2372       EGF-LIKE 40, CALCIUM-BINDING.
FT   REPEAT      2373   2441       TGFBP 7.
FT   DOMAIN      2442   2483       EGF-LIKE 41, CALCIUM-BINDING.
FT   DOMAIN      2484   2524       EGF-LIKE 42, CALCIUM-BINDING.
FT   DOMAIN      2525   2563       EGF-LIKE 43, CALCIUM-BINDING.
FT   DOMAIN      2564   2606       EGF-LIKE 44, CALCIUM-BINDING.
FT   DOMAIN      2607   2646       EGF-LIKE 45, CALCIUM-BINDING.
FT   DOMAIN      2647   2687       EGF-LIKE 46, CALCIUM-BINDING.
FT   DOMAIN      2688   2727       EGF-LIKE 47, CALCIUM-BINDING.
FT   DISULFID     115    124       BY SIMILARITY.
FT   DISULFID     119    130       BY SIMILARITY.
FT   DISULFID     132    141       BY SIMILARITY.
FT   DISULFID     149    159       BY SIMILARITY.
FT   DISULFID     153    164       BY SIMILARITY.
FT   DISULFID     166    175       BY SIMILARITY.
FT   DISULFID     180    190       BY SIMILARITY.
FT   DISULFID     184    196       BY SIMILARITY.
FT   DISULFID     198    207       BY SIMILARITY.
FT   DISULFID     280    292       BY SIMILARITY.
FT   DISULFID     287    301       BY SIMILARITY.
FT   DISULFID     303    316       BY SIMILARITY.
FT   DISULFID     322    334       BY SIMILARITY.
FT   DISULFID     329    343       BY SIMILARITY.
FT   DISULFID     345    358       BY SIMILARITY.
FT   DISULFID     491    503       BY SIMILARITY.
FT   DISULFID     498    512       BY SIMILARITY.
FT   DISULFID     514    526       BY SIMILARITY.
FT   DISULFID     532    542       BY SIMILARITY.
FT   DISULFID     537    551       BY SIMILARITY.
FT   DISULFID     553    566       BY SIMILARITY.
FT   DISULFID     572    584       BY SIMILARITY.
FT   DISULFID     579    593       BY SIMILARITY.
FT   DISULFID     595    608       BY SIMILARITY.
FT   DISULFID     614    625       BY SIMILARITY.
FT   DISULFID     620    634       BY SIMILARITY.
FT   DISULFID     636    649       BY SIMILARITY.
FT   DISULFID     655    666       BY SIMILARITY.
FT   DISULFID     661    675       BY SIMILARITY.
FT   DISULFID     677    690       BY SIMILARITY.
FT   DISULFID     765    777       BY SIMILARITY.
FT   DISULFID     772    786       BY SIMILARITY.
FT   DISULFID     788    801       BY SIMILARITY.
FT   DISULFID     807    819       BY SIMILARITY.
FT   DISULFID     814    828       BY SIMILARITY.
FT   DISULFID     830    843       BY SIMILARITY.
FT   DISULFID     849    859       BY SIMILARITY.
FT   DISULFID     854    868       BY SIMILARITY.
FT   DISULFID     870    883       BY SIMILARITY.
FT   DISULFID     952    964       BY SIMILARITY.
FT   DISULFID     959    973       BY SIMILARITY.
FT   DISULFID     975    988       BY SIMILARITY.
FT   DISULFID    1070   1082       BY SIMILARITY.
FT   DISULFID    1077   1091       BY SIMILARITY.
FT   DISULFID    1093   1106       BY SIMILARITY.
FT   DISULFID    1112   1124       BY SIMILARITY.
FT   DISULFID    1119   1133       BY SIMILARITY.
FT   DISULFID    1135   1149       BY SIMILARITY.
FT   DISULFID    1155   1167       BY SIMILARITY.
FT   DISULFID    1162   1176       BY SIMILARITY.
FT   DISULFID    1178   1191       BY SIMILARITY.
FT   DISULFID    1197   1209       BY SIMILARITY.
FT   DISULFID    1204   1218       BY SIMILARITY.
FT   DISULFID    1220   1233       BY SIMILARITY.
FT   DISULFID    1239   1250       BY SIMILARITY.
FT   DISULFID    1246   1259       BY SIMILARITY.
FT   DISULFID    1261   1274       BY SIMILARITY.
FT   DISULFID    1280   1292       BY SIMILARITY.
FT   DISULFID    1287   1301       BY SIMILARITY.
FT   DISULFID    1303   1316       BY SIMILARITY.
FT   DISULFID    1322   1334       BY SIMILARITY.
FT   DISULFID    1329   1343       BY SIMILARITY.
FT   DISULFID    1345   1358       BY SIMILARITY.
FT   DISULFID    1364   1377       BY SIMILARITY.
FT   DISULFID    1371   1386       BY SIMILARITY.
FT   DISULFID    1388   1399       BY SIMILARITY.
FT   DISULFID    1405   1418       BY SIMILARITY.
FT   DISULFID    1412   1427       BY SIMILARITY.
FT   DISULFID    1429   1440       BY SIMILARITY.
FT   DISULFID    1446   1458       BY SIMILARITY.
FT   DISULFID    1453   1467       BY SIMILARITY.
FT   DISULFID    1469   1482       BY SIMILARITY.
FT   DISULFID    1488   1499       BY SIMILARITY.
FT   DISULFID    1494   1508       BY SIMILARITY.
FT   DISULFID    1510   1523       BY SIMILARITY.
FT   DISULFID    1529   1540       BY SIMILARITY.
FT   DISULFID    1535   1549       BY SIMILARITY.
FT   DISULFID    1551   1564       BY SIMILARITY.
FT   DISULFID    1647   1659       BY SIMILARITY.
FT   DISULFID    1654   1668       BY SIMILARITY.
FT   DISULFID    1670   1683       BY SIMILARITY.
FT   DISULFID    1689   1701       BY SIMILARITY.

  Query Match             12.2%;  Score 183.5;  DB 1;  Length 2907;
  Best Local Similarity   23.9%;  Pred. No. 2.4e-06;
  Matches   60;  Conservative   19;  Mismatches   85;  Indels   87;  Gaps    7;

Qy          2 RGSQEVLLMWLLVLAVGGTEHAY-------RPGRRVCAVRAHGDPVSESFVQRVYQPFLT 54
              || ||:|         |   |:|        ||   | |    :   : |  |   | :
Db         77 RGQQEILRG---PNVCGSRFHSYCCPGWKTLPGGNQCIVPICRNSCGDGFCSR---PNMC 130

Qy         55 TCDGHRACST--YRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCR 112
              ||   :   |   ::| : : |   |   |     |  |:  |      ||  :|:  |:
Db        131 TCSSGQISPTCGRKSIQQCSVRCMNGGTCADDHCQCQKGYIGT-----YCGQPVCENGCQ 185

Qy        113 NGGSCVQPGRCRCPAGWRGDTCQ------------------------------------- 135
              ||| |: | || |  |: |  |:
Db        186 NGGRCIGPNRCACVYGFTGPQCERDYRTGPCFTQVNNQMCQGQLTGIVCTKTLCCATIGR 245

Qy        136 -----------------------------SDVDECSARRGGCP-QRCINTAGSYWCQCWE 165
```

```
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 2.
DR   SMART; SM00179; EGF_CA; 3.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50026; EGF_3; 6.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC_1; 2.
DR   PROSITE; PS50184; VWFC_2; 2.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    816       PROTEIN KINASE C-BINDING PROTEIN NELL2.
FT   DOMAIN       30    258       TSP N-TERMINAL.
FT   DOMAIN      272    331       VWFC 1.
FT   DOMAIN      332    396       VWFC 2.
FT   DOMAIN      397    439       EGF-LIKE 1.
FT   DOMAIN      440    481       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      482    522       EGF-LIKE 3.
FT   DOMAIN      521    553       EGF-LIKE 4.
FT   DOMAIN      555    601       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      602    637       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      638    693       VWFC 3.
FT   DOMAIN      698    756       VWFC 4.
FT   DOMAIN      758    813       VWFC 5.
FT   DISULFID    401    413       BY SIMILARITY.
FT   DISULFID    407    422       BY SIMILARITY.
FT   DISULFID    424    438       BY SIMILARITY.
FT   DISULFID    444    457       BY SIMILARITY.
FT   DISULFID    451    466       BY SIMILARITY.
FT   DISULFID    468    480       BY SIMILARITY.
FT   DISULFID    486    499       BY SIMILARITY.
FT   DISULFID    493    508       BY SIMILARITY.
FT   DISULFID    510    521       BY SIMILARITY.
FT   DISULFID    525    535       BY SIMILARITY.
FT   DISULFID    529    541       BY SIMILARITY.
FT   DISULFID    543    552       BY SIMILARITY.
FT   DISULFID    559    572       BY SIMILARITY.
FT   DISULFID    566    581       BY SIMILARITY.
FT   DISULFID    583    600       BY SIMILARITY.
FT   DISULFID    606    619       BY SIMILARITY.
FT   DISULFID    613    628       BY SIMILARITY.
FT   DISULFID    630    636       BY SIMILARITY.
FT   CARBOHYD     53     53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    298    298       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    635    635       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   816 AA;  91163 MW;  5BDD0A946F87E74D CRC64;

  Query Match             12.3%;  Score 185;  DB 1;  Length 816;
  Best Local Similarity   29.4%;  Pred. No. 4.8e-07;
  Matches   47;  Conservative   21;  Mismatches   64;  Indels   28;  Gaps    7;

Qy          26 PGRRVCAVRAHGDPVSESFVQRVYQPFLTT---CDGHRACSTYRTIYRTAYRRSPGLAPA 82
               ||  :|  :  |    : :    :   |||   || :  |  : |:
Db         461 PGSFMCVCKT-GYIRIDDYSCTEHDECLTTQHNCDENALC--FNTV------------G 504

Qy          83 RPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECS 142
                    | ||:       |    | |:  |||||:|:    | || |: | :|::|:||||
Db         505 GHNCVCKPGYTGN----GTTCKAFCKDGCRNGGACIAANVCACPQGFTGPSCETDIDECS 560

Qy         143 ARRGGCPQR--CINTAGSYWCQCWEGHSLSADGTLCVPKG 180
                    |  |  |||  | | |:| :|:    |  :  | |
Db         561 EGFVQCDSRANCINLPGWYHCECRDGYH---DNGMFAPGG 597


RESULT 4
```

```
FBN2_MOUSE
ID   FBN2_MOUSE     STANDARD;      PRT;  2907 AA.
AC   Q61555; Q63957;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Fibrillin 2 precursor.
GN   FBN2 OR FBN-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95263670; PubMed=7744963;
RA   Zhang H., Hu W., Ramirez F.;
RT   "Developmental expression of fibrillin genes suggests heterogeneity
RT   of extracellular microfibrils.";
RL   J. Cell Biol. 129:1165-1176(1995).
RN   [2]
RP   SEQUENCE OF 210-317 FROM N.A.
RX   MEDLINE=94140368; PubMed=8307578;
RA   Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,
RA   Francke U.;
RT   "Fibrillin genes map to regions of conserved mouse/human synteny on
RT   mouse chromosomes 2 and 18.";
RL   Genomics 18:667-672(1993).
CC   -!- FUNCTION: Structural component of connective tissue microfibrils
CC       that binds calcium. Fibrillin-2-containing microfibrils regulate
CC       the early process of elastic fiber assembly.
CC   -!- SIMILARITY: Contains 47 EGF-like domains.
CC   -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L39790; AAA74908.1; -.
DR   EMBL; S69359; AAC60685.1; -.
DR   PIR; A57278; A57278.
DR   HSSP; P35555; 1EMN.
DR   MGD; MGI:95490; Fbn2.
DR   GO; GO:0030326; P:limb morphogenesis; IMP.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 44.
DR   Pfam; PF00683; TB; 9.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00179; EGF_CA; 43.
DR   PROSITE; PS00010; ASX_HYDROXYL; 43.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 36.
DR   PROSITE; PS50026; EGF_3; 45.
DR   PROSITE; PS01187; EGF_CA; 43.
KW   Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW   Repeat; Signal; Multigene family.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   2907       FIBRILLIN 2.
FT   DOMAIN      111    142       EGF-LIKE 1.
FT   DOMAIN      145    176       EGF-LIKE 2.
FT   DOMAIN      176    208       EGF-LIKE 3.
FT   DOMAIN      276    317       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      318    359       EGF-LIKE 5, CALCIUM-BINDING.
FT   REPEAT      360    426       TGFBP 1.
FT   DOMAIN      487    527       EGF-LIKE 6.
FT   DOMAIN      528    567       EGF-LIKE 7, CALCIUM-BINDING.
```

```
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00179; EGF_CA; 11.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS50026; EGF_3; 28.
DR   PROSITE; PS01187; EGF_CA; 9.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
KW   Triplet repeat expansion; Alternative splicing.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   2003       NEUROGENIC LOCUS NOTCH PROTEIN HOMOLOG 4.
FT   CHAIN      1432   2003       NOTCH EXTRACELLULAR TRUNCATION
FT                                (BY SIMILARITY).
FT   CHAIN      1467   2003       NOTCH INTRACELLULAR DOMAIN
FT                                (BY SIMILARITY).
FT   DOMAIN       24   1447       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1448   1468       POTENTIAL.
FT   DOMAIN     1469   2003       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       24     63       EGF-LIKE 1.
FT   DOMAIN       64    115       EGF-LIKE 2.
FT   DOMAIN      118    155       EGF-LIKE 3.
FT   DOMAIN      156    192       EGF-LIKE 4.
FT   DOMAIN      194    232       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      234    274       EGF-LIKE 6.
FT   DOMAIN      276    312       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      314    353       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      355    391       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      392    430       EGF-LIKE 10.
FT   DOMAIN      432    473       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      475    511       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      513    549       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      551    587       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      589    625       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      626    659       EGF-LIKE 16.
FT   DOMAIN      661    689       EGF-LIKE 17.
FT   DOMAIN      691    727       EGF-LIKE 18.
FT   DOMAIN      729    765       EGF-LIKE 19.
FT   DOMAIN      767    803       EGF-LIKE 20.
FT   DOMAIN      806    842       EGF-LIKE 21.
FT   DOMAIN      844    880       EGF-LIKE 22.
FT   DOMAIN      882    928       EGF-LIKE 23.
FT   DOMAIN      930    966       EGF-LIKE 24.
FT   DOMAIN      968   1004       EGF-LIKE 25.
FT   DOMAIN     1006   1044       EGF-LIKE 26.
FT   DOMAIN     1046   1085       EGF-LIKE 27.
FT   DOMAIN     1087   1126       EGF-LIKE 28.
FT   DOMAIN     1130   1171       EGF-LIKE 29.
FT   DOMAIN     1472   1476       POLY-ARG.
FT   REPEAT     1165   1212       LIN/NOTCH 1.
FT   REPEAT     1213   1246       LIN/NOTCH 2.
FT   REPEAT     1247   1286       LIN/NOTCH 3.
FT   REPEAT     1633   1665       ANK 1.
FT   REPEAT     1666   1698       ANK 2.
FT   REPEAT     1700   1732       ANK 3.
FT   REPEAT     1733   1765       ANK 4.
FT   REPEAT     1766   1798       ANK 5.
FT   DISULFID     28     41       BY SIMILARITY.
FT   DISULFID     35     51       BY SIMILARITY.
FT   DISULFID     53     62       BY SIMILARITY.
FT   DISULFID     68     80       BY SIMILARITY.
FT   DISULFID     74    103       BY SIMILARITY.
FT   DISULFID    105    114       BY SIMILARITY.
FT   DISULFID    122    133       BY SIMILARITY.
FT   DISULFID    127    143       BY SIMILARITY.
FT   DISULFID    145    154       BY SIMILARITY.
FT   DISULFID    160    171       BY SIMILARITY.
FT   DISULFID    165    180       BY SIMILARITY.
FT   DISULFID    182    191       BY SIMILARITY.
```

```
FT   DISULFID    198    211       BY SIMILARITY.

  Query Match             12.5%;  Score 188.5;  DB 1;  Length 2003;
  Best Local Similarity   29.6%;  Pred. No. 6.7e-07;
  Matches   61;  Conservative   17;  Mismatches   87;  Indels   41;  Gaps   11;

Qy         80 APARPRYACCPGWK-RTSGLPGACGAAICQPPCRNGGSCVQPG---RCRCPAGWRGDTCQ 135
              |  ||: :| |||      |   | |     || ||| |:      :| || |: |  |:
Db        137 ASGRPQCSCMPGWTGEQCQLRDFCSA----NPCVNGGVCLATYPQIQCHCPPGFEGHACE 192

Qy        136 SDVDECSARRGGCPQ--RCINTAGSYWCQCWEGHSLSADGTLCVPKGG--PPRVAPNPTG 191
               ||:||    | ||:   | || ||: | |  |     :|  |  : |  |||   |  |
Db        193 RDVNECFQDPGPCPKGTSCHNTLGSFQCLCPVGQ----EGPRCELRAGPCPPRGCSN-GG 247

Qy        192 VDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPD----PGSLLVHSFQQL 247
                  | |:           :    | | |   :         ||    | : : |  |
Db        248 TCQLMPEK-----------DSTFHLCLCPPGFIG--------PDCEVNPDNCVSHQCQNG 288

Qy        248 GRI-DSLSEQISFLEEQLGSCSCKKD 272
              |   | |        |      | :|
Db        289 GTCQDGLDTYTCLCPETWTGWDCSED 314

RESULT 3
NEL2_MOUSE
ID   NEL2_MOUSE     STANDARD;      PRT;   816 AA.
AC   Q61220;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein kinase C-binding protein NELL2 precursor (NEL-like protein 2)
DE   (MEL91 protein).
GN   NELL2 OR MEL91.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Elkins D.A., Rossi J.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   TSP N-TERMINAL DOMAIN.
RX   MEDLINE=98153258; PubMed=9480764;
RA   Beckmann G., Hanke J., Bork P., Reich J.;
RT   "Merging extracellular domains: fold prediction for laminin G-like
RT   and amino-terminal thrombospondin-like modules based on homology to
RT   pentraxins.";
RL   J. Mol. Biol. 275:725-730(1998).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC   -!- SIMILARITY: Contains 5 VWFC domains.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U59230; AAB02924.1; ALT_INIT.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:1858510; Nell2.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR008985; ConA_like_lec_gl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00008; EGF; 4.
```

```
  Query Match             13.0%;  Score 196;  DB 1;  Length 1964;
  Best Local Similarity   34.2%;  Pred. No. 1.7e-07;
  Matches   50;  Conservative   12;  Mismatches   54;  Indels   30;  Gaps    7;

Qy         80 APARPRYACCPGWK-RTSGLPGACGAAICQPPCRNGGSCVQPG---RCRCPAGWRGDTCQ 135
              |  ||: :| |||      |   | |     || ||| |:      :|||| |: | ||:
Db        134 ASGRPQCSCEPGWTGEQCQLRDFCSA----NPCANGGVCLATYPQIQCRCPPGFEGHTCE 189

Qy        136 SDVDECSARRGGCPQ--RCINTAGSYWCQC---WEGHSLSADGTLCVP----KGGPPRVA 186
               |::||    | |||   | || ||| | |    ||         | |     || ::
Db        190 RDINECFLEPGPCPQGTSCHNTLGSYQCLCPVGQEGPQCKLRKGACPPGSCLNGGTCQLV 249

Qy        187 PNP-------------TGVDSAMKEE 199
              |               ||:|  |  :
Db        250 PEGHSTFHLCLCPPGFTGLDCEMNPD 275


RESULT 2
NTC4_HUMAN
ID   NTC4_HUMAN     STANDARD;      PRT;  2003 AA.
AC   Q99466; O00306; Q99458; Q99940; Q9H3S8; Q9UII9; Q9UIJ0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE   (hNotch4).
GN   NOTCH4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
RC   TISSUE=Placenta;
RX   MEDLINE=97311416; PubMed=9168133;
RA   Sugaya K., Sasanuma S.-I., Nohata J., Kimura T., Fukagawa T.,
RA   Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
RT   "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
RT   human counterpart gene of mouse proto-oncogene Int3.";
RL   Gene 189:235-244(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Bone marrow, and Heart;
RX   MEDLINE=98360091; PubMed=9693032;
RA   Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
RA   Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
RT   "Cloning, characterization, and the complete 56.8-kilobase DNA
RT   sequence of the human NOTCH4 gene.";
RL   Genomics 51:45-58(1998).
RN   [3]
RP   SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
RA   Miyagawa T., Tokunaga K., Hojho H.;
RT   "Human notch4 gene variant.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   IDENTIFICATION OF LIGANDS.
RX   MEDLINE=99180765; PubMed=10079256;
RA   Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA   Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT   "Human ligands of the Notch receptor.";
RL   Am. J. Pathol. 154:785-794(1999).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs. May regulate branching morphogenesis in the
CC       developing vascular system (By similarity).
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=1;
CC         IsoId=Q99466-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q99466-2; Sequence=VSP_001406;
CC       Name=3;
CC         IsoId=Q99466-3; Sequence=VSP_001407;
CC   -!- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
CC       the lung and placenta and at low levels in the liver, skeletal
CC       muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
CC       and fetal liver. No expression was seen in adult brain or
CC       peripheral blood leukocytes.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane (By similarity).
CC   -!- PTM: Phosphorylated (By similarity).
CC   -!- POLYMORPHISM: The poly-Leu region of NOTCH4 (in the signal
CC       peptide) is polymorphic and the number of Leu varies in the
CC       population (from 6 to 12).
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 28 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC       in position 1438 to 1463.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D63395; BAA09708.1; ALT_FRAME.
DR   EMBL; D86566; BAA13116.1; -.
DR   EMBL; U95299; AAC32288.1; -.
DR   EMBL; U89335; AAC63097.1; -.
DR   EMBL; AB023961; BAB20317.1; -.
DR   EMBL; AB024520; BAA88951.1; -.
DR   EMBL; AB024578; BAA88952.1; -.
DR   HSSP; P08709; 1BF9.
DR   Genew; HGNC:7884; NOTCH4.
DR   MIM; 164951; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 26.
DR   Pfam; PF00066; notch; 2.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
```

```
RT   Notch4 in embryonic endothelium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
RN   [7]
RP   SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
RP   OF VAL-1463.
RX   MEDLINE=21523956; PubMed=11518718;
RA   Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT   "Murine notch homologs (N1-4) undergo presenilin-dependent
RT   proteolysis.";
RL   J. Biol. Chem. 276:40268-40273(2001).
RN   [8]
RP   POST-TRANSLATIONAL PROCESSING.
RX   MEDLINE=21374376; PubMed=11459941;
RA   Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT   "Conservation of the biochemical mechanisms of signal transduction
RT   among mammalian Notch family members.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC   -!- FUNCTION: Functions as a receptor for membrane-bound ligands
CC       Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
CC       Upon ligand activation through the released notch intracellular
CC       domain (NICD) it forms a transcriptional activator complex with
CC       RBP-J kappa and activates genes of the enhancer of split locus.
CC       Affects the implementation of differentiation, proliferation and
CC       apoptotic programs (By similarity). May regulate branching
CC       morphogenesis in the developing vascular system.
CC   -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
CC       terminal fragment N(EC) which are probably linked by disulfide
CC       bonds.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC       proteolytical processing NICD is translocated to the nucleus.
CC   -!- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
CC       kidney, and at lower levels in the ovary and skeletal muscle. A
CC       very low expression is seen in the brain, intestine, liver and
CC       testis.
CC   -!- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
CC       embryonic development from 9.0 dpc.
CC   -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC       which is proteolytically cleaved by a furin-like convertase in the
CC       trans-Golgi network before it reaches the plasma membrane to yield
CC       an active, ligand-accessible form. Cleavage results in a C-
CC       terminal fragment N(TM) and a N-terminal fragment N(EC). Following
CC       ligand binding, it is cleaved by TNF-alpha converting enzyme
CC       (TACE) to yield a membrane-associated intermediate fragment called
CC       notch extracellular truncation (NEXT). This fragment is then
CC       cleaved by presenilin dependent gamma-secretase to release a
CC       notch-derived peptide containing the intracellular domain (NICD)
CC       from the membrane.
CC   -!- PTM: Phosphorylated.
CC   -!- DISEASE: Loss of the extracellular domain causes constitutive
CC       activation of the Notch protein, which leads to hyperproliferation
CC       of glandular epithelial tissues and development of mammary
CC       carcinomas.
CC   -!- SIMILARITY: Belongs to the NOTCH family.
CC   -!- SIMILARITY: Contains 29 EGF-like domains.
CC   -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC   -!- SIMILARITY: Contains 5 ANK repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M80456; AAB38377.1; -.
DR   EMBL; U43691; AAC52630.1; -.
DR   EMBL; U43691; AAC52631.1; -.
DR   EMBL; AF030001; AAB82004.1; -.
DR   EMBL; AB016771; BAA32281.1; ALT_SEQ.
DR   EMBL; AB016772; BAA32283.1; ALT_INIT.
DR   EMBL; AB016773; BAA32284.1; ALT_INIT.
DR   EMBL; AB016774; BAA32285.1; -.
```

```
DR   PIR; A38072; TVMVT3.
DR   PIR; T09059; T09059.
DR   HSSP; P08709; 1BF9.
DR   MGD; MGI:107471; Notch4.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008297; Notch.
DR   InterPro; IPR000800; Notch_dom.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 27.
DR   Pfam; PF00066; notch; 2.
DR   PIRSF; PIRSF002279; Notch; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 6.
DR   SMART; SM00179; EGF_CA; 11.
DR   SMART; SM00004; NL; 2.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 5.
DR   PROSITE; PS00010; ASX_HYDROXYL; 11.
DR   PROSITE; PS00022; EGF_1; 28.
DR   PROSITE; PS01186; EGF_2; 21.
DR   PROSITE; PS50026; EGF_3; 27.
DR   PROSITE; PS01187; EGF_CA; 9.
KW   Receptor; Transcription regulation; Activator; Differentiation;
KW   Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW   Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1964       NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT   CHAIN      1411   1964       TRANSFORMING PROTEIN INT-3.
FT   CHAIN      1428   1964       NOTCH EXTRACELLULAR TRUNCATION.
FT   CHAIN      1463   1964       NOTCH INTRACELLULAR DOMAIN.
FT   DOMAIN       21   1443       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1444   1464       POTENTIAL.
FT   DOMAIN     1465   1964       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       21     60       EGF-LIKE 1.
FT   DOMAIN       61    112       EGF-LIKE 2.
FT   DOMAIN      115    152       EGF-LIKE 3.
FT   DOMAIN      153    189       EGF-LIKE 4.
FT   DOMAIN      191    229       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      231    271       EGF-LIKE 6.
FT   DOMAIN      273    309       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      311    350       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      352    388       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      389    427       EGF-LIKE 10.
FT   DOMAIN      429    470       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      472    508       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      510    546       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      548    584       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      586    622       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      623    656       EGF-LIKE 16.
FT   DOMAIN      658    686       EGF-LIKE 17.
FT   DOMAIN      688    724       EGF-LIKE 18.
FT   DOMAIN      726    762       EGF-LIKE 19.
FT   DOMAIN      764    800       EGF-LIKE 20.
FT   DOMAIN      803    839       EGF-LIKE 21.
FT   DOMAIN      841    877       EGF-LIKE 22.
FT   DOMAIN      878    924       EGF-LIKE 23.
FT   DOMAIN      926    962       EGF-LIKE 24.
FT   DOMAIN      964   1000       EGF-LIKE 25.
FT   DOMAIN     1002   1040       EGF-LIKE 26.
FT   DOMAIN     1042   1081       EGF-LIKE 27.
FT   DOMAIN     1083   1122       EGF-LIKE 28.
FT   DOMAIN     1126   1167       EGF-LIKE 29.
FT   REPEAT     1168   1208       LIN/NOTCH 1.
FT   REPEAT     1209   1242       LIN/NOTCH 2.
FT   REPEAT     1243   1282       LIN/NOTCH 3.
```

```
                      GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         April 22, 2004, 12:42:05 ; Search time 18 Seconds
                                          (without alignments)
                                          789.731 Million cell updates/sec

Title:          US-09-978-191A-506
Perfect score:  1505
Sequence:       1 MRGSQEVLLMWLLVLAVGGT..........SEQISFLEEQLGSCSCKKDS 273

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:      141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_42:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                               SUMMARIES
                 %
Result         Query
  No.   Score  Match Length DB  ID                                Description
 ---------------------------------------------------------------------------
     1    196   13.0   1964  1  NTC4_MOUSE                        P31695 mus musculu
     2  188.5   12.5   2003  1  NTC4_HUMAN                        Q99466 homo sapien
     3    185   12.3    816  1  NEL2_MOUSE                        Q61220 mus musculu
     4  183.5   12.2   2907  1  FBN2_MOUSE                        Q61555 mus musculu
     5  182.5   12.1    816  1  NEL2_HUMAN                        Q99435 homo sapien
     6    182   12.1    810  1  NEL1_HUMAN                        Q92832 homo sapien
     7  180.5   12.0    816  1  NEL2_RAT                          Q62918 rattus norv
     8    180   12.0    810  1  NEL1_RAT                          Q62919 rattus norv
     9  176.5   11.7    816  1  NEL_CHICK                         Q90827 gallus gall
    10  176.5   11.7   2871  1  FBN1_HUMAN                        P35555 homo sapien
    11  175.5   11.7   2321  1  NTC3_HUMAN                        Q9um47 homo sapien
    12    175   11.6    652  1  CD93_HUMAN                        Q9npy3 homo sapien
    13    175   11.6   2703  1  NOTC_DROME                        P07207 drosophila
    14  173.5   11.5   1247  1  NIDO_HUMAN                        P14543 homo sapien
    15  172.5   11.5   2871  1  FBN1_MOUSE                        Q61554 mus musculu
    16    172   11.4   2531  1  NTC1_MOUSE                        Q01705 mus musculu
    17  171.5   11.4   2871  1  FBN1_BOVIN                        P98133 bos taurus
    18  170.5   11.3   2871  1  FBN1_PIG                          Q9tv36 sus scrofa
    19    170   11.3    675  1  PRTS_MOUSE                        Q08761 mus musculu
    20    170   11.3   2911  1  FBN2_HUMAN                        P35556 homo sapien
    21  169.5   11.3   2318  1  NTC3_MOUSE                        Q61982 mus musculu
    22    169   11.2    833  1  DL_DROME                          P10041 drosophila
    23    168   11.2    675  1  PRTS_RAT                          P53813 rattus norv
    24    165   11.0   1238  1  JAG2_HUMAN                        Q9y219 homo sapien
    25    165   11.0   1247  1  JAG2_MOUSE                        Q9qye5 mus musculu
    26    164   10.9   1202  1  JAG2_RAT                          P97607 rattus norv
    27    164   10.9   1429  1  LI12_CAEEL                        P14585 caenorhabdi
    28    164   10.9   2524  1  NOTC_XENLA                        P21783 xenopus lae
    29  163.5   10.9   2556  1  NTC1_HUMAN                        P46531 homo sapien
    30    163   10.8    644  1  CD93_MOUSE                        O89103 mus musculu
    31  162.5   10.8   2319  1  NTC3_RAT                          Q9r172 rattus norv
    32    162   10.8    459  1  PRTC_PIG                          Q9glp2 sus scrofa
    33    162   10.8    646  1  PRTS_RABIT                        P98118 oryctolagus
```

```
    34    162   10.8    833  1  SRC2_MOUSE                        P59222 mus musculu
    35    162   10.8    870  1  SRC2_HUMAN                        Q96gp6 homo sapien
    36    161   10.7    675  1  PRTS_BOVIN                        P07224 bos taurus
    37  160.5   10.7    618  1  DLL3_HUMAN                        Q9nyj7 homo sapien
    38  160.5   10.7   2531  1  NTC1_RAT                          Q07008 rattus norv
    39    160   10.6   2437  1  NTC1_BRARE                        P46530 brachydanio
    40  159.5   10.6    643  1  CD93_RAT                          Q9et61 rattus norv
    41  158.5   10.5    407  1  FA7_BOVIN                         P22457 bos taurus
    42  158.5   10.5    456  1  PRTC_BOVIN                        P00745 bos taurus
    43  157.5   10.5    379  1  WIF1_MOUSE                        Q9wua1 mus musculu
    44    157   10.4    676  1  PRTS_HUMAN                        P07225 homo sapien
    45  156.5   10.4   1213  1  JAG3_BRARE                        Q90y54 brachydanio

                                ALIGNMENTS

RESULT 1
NTC4_MOUSE
ID   NTC4_MOUSE     STANDARD;      PRT;  1964 AA.
AC   P31695; O35442; O88314; O88316; Q62389; Q62390; Q9R1W9; Q9R1X0;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE   [Contains: Transforming protein Int-3].
GN   NOTCH4 OR INT3 OR INT-3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92194507; PubMed=1312643;
RA   Robbins J., Blondel B.J., Gallahan D., Callahan R.;
RT   "Mouse mammary tumor gene int-3: a member of the notch gene family
RT   transforms mammary epithelial cells.";
RL   J. Virol. 66:2594-2599(1992).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RX   MEDLINE=97294599; PubMed=9150355;
RA   Gallahan D., Callahan R.;
RT   "The mouse mammary tumor associated gene INT3 is a unique member of
RT   the NOTCH gene family (NOTCH4).";
RL   Oncogene 14:1883-1890(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung, and Testis;
RX   MEDLINE=96281668; PubMed=8681805;
RA   Uyttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT   "Notch4/int-3, a mammary proto-oncogene, is an endothelial
RT   cell-specific mammalian Notch gene.";
RL   Development 122:2251-2259(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA   Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT   "Sequence of the mouse major histocompatibility locus class III
RT   region.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1436-1600 FROM N.A.
RX   MEDLINE=99252212; PubMed=10233982;
RA   Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT   "Intracisternal type A particle-mediated activation of the Notch4/int3
RT   gene in a mouse mammary tumor: generation of truncated Notch4/int3
RT   mRNAs by retroviral splicing events.";
RL   J. Virol. 73:5166-5171(1999).
RN   [6]
RP   FUNCTION.
RX   MEDLINE=21244657; PubMed=11344305;
RA   Uyttendaele H., Ho J., Rossant J., Kitajewski J.;
RT   "Vascular patterning defects associated with expression of activated
```

```
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR000998; MAM_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00181; EGF; 5.
DR   SMART; SM00179; EGF_CA; 3.
DR   SMART; SM00137; MAM; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS50060; MAM_2; 1.
SQ   SEQUENCE   553 AA;  61317 MW;  3AE93A0362E861E0 CRC64;

  Query Match             16.7%;  Score 251.5;  DB 4;  Length 553;
  Best Local Similarity   36.8%;  Pred. No. 3.1e-14;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy         58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
              |: | : :  :  :|  | ||:     : ||| ||:| |   | |  | |:| |:  | |
Db         21 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 74

Qy        118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
              | | :|||  |: | ||  |||:||  :   |  ||:|| ||| | |  || |  | | ||
Db         75 VGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 133

Qy        178 ----------------PKGGPP--------RVAPN 188
                               : ||         |:|||
Db        134 NSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168


RESULT 15
Q9NY67
ID   Q9NY67      PRELIMINARY;      PRT;   554 AA.
AC   Q9NY67;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   W80.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Teratocarcinoma, and Neuron;
RA   Franco B.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Teratocarcinoma, and Neuron;
RX   MEDLINE=20241927; PubMed=10777661;
RA   Buchner G., Orfanelli U., Quaderi N., Bassi M.T., Andolfi G.;
RT   "Identification of a new EGF-repeat-containing gene from human Xp22:
RT   Acandidate for developmental disorders.";
RL   Genomics 65:16-23(2000).
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AJ245671; CAB92132.1; -.
DR   HSSP; P00736; 1APQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000998; MAM_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00179; EGF_CA; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Hypothetical protein; EGF-like domain; Glycoprotein.
SQ   SEQUENCE   554 AA;  61388 MW;  D519238F2A604101 CRC64;

  Query Match             16.7%;  Score 251.5;  DB 4;  Length 554;
  Best Local Similarity   36.8%;  Pred. No. 3.1e-14;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy         58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
              |: | : :  :  :|  | ||:     : ||| ||:| |   | |  | |:| |:  | |
Db         21 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 74

Qy        118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
              | | :|||  |: | ||  |||:||  :   |  ||:|| ||| | |  || |  | | ||
Db         75 VGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 133

Qy        178 ----------------PKGGPP--------RVAPN 188
                               : ||         |:|||
Db        134 NSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168


Search completed: April 22, 2004, 12:46:49
Job time : 47 secs
```

```
Db         43 VCAEQKLTLVGHRQPCVQAFSRIVPVWRRTGCAQQAWCIGQERRTVYYMSYRQVYA-TEA 101

Qy         83 RPRYACCPGWKRTSGLPGA------CGAAI--CQPPCRN--GGSCVQPGRCRCPAGWR-- 130
              |  : ||||| :  |  |       | :|   |: || |  ||       |||| |::
Db        102 RTVFRCCPGWSQKPGQEGCLSDVDECASANGGCEGPCCNTVGGF-----YCRCPPGYQLQ 156

Qy        131 --GDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
                | ||| ||||| |  |||  ||:|| ||| |:|  |  |  ||  |:
Db        157 GDGKTCQ-DVDECRAHNGGCQHRCVNTPGSYLCECKPGFRLHTDGRTCL 204


RESULT 12
Q9UFK6
ID   Q9UFK6      PRELIMINARY;      PRT;   558 AA.
AC   Q9UFK6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP564P2063.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Duesterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AL117610; CAB56014.1; -.
DR   PIR; T17324; T17324.
DR   HSSP; P00736; 1APQ.
DR   Genew; HGNC:3235; EGFL6.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005178; F:integrin binding; TAS.
DR   GO; GO:0007049; P:cell cycle; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000998; MAM_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00179; EGF_CA; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50060; MAM_2; 1.
KW   Hypothetical protein; EGF-like domain; Glycoprotein.
FT   NON_TER       1      1
SQ   SEQUENCE   558 AA;  61828 MW;  AA38D7DCE402BFA3 CRC64;

  Query Match             16.8%;  Score 252.5;  DB 4;  Length 558;
  Best Local Similarity   36.8%;  Pred. No. 2.6e-14;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy         58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
              |: | : :  :  :|  | ||:     : ||| ||:| |   | |  | |:| |:  | |
Db         26 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 79

Qy        118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
              | | :|||  |: | ||  ||:||  :   |  ||:|| ||| | |  || |  | | ||
Db         80 VGPNKCRCLPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 138

Qy        178 ----------------PKGGPP--------RVAPN 188
                               : ||         |:|||
Db        139 YSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 173


RESULT 13
Q9NZL7
ID   Q9NZL7      PRELIMINARY;      PRT;   553 AA.
```

```
AC   Q9NZL7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Epidermal growth factor repeat containing protein.
GN   EGFL6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20079166; PubMed=10610727;
RA   Yeung G., Mulero J.J., Berntsen R.P., Loeb D.B., Drmanac R.,
RA   Ford J.E.;
RT   "Cloning of a novel epidermal growth factor repeat containing gene
RT   EGFL6: expressed in tumor and fetal tissues.";
RL   Genomics 62:304-307(1999).
CC   -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR   EMBL; AF186084; AAF27812.1; -.
DR   HSSP; P00736; 1APQ.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000998; MAM_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00629; MAM; 1.
DR   SMART; SM00179; EGF_CA; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS50060; MAM_2; 1.
KW   EGF-like domain; Glycoprotein.
SQ   SEQUENCE   553 AA;  61314 MW;  2FF55F167857DE50 CRC64;

  Query Match             16.7%;  Score 251.5;  DB 4;  Length 553;
  Best Local Similarity   36.8%;  Pred. No. 3.1e-14;
  Matches   57;  Conservative   17;  Mismatches   50;  Indels   31;  Gaps    7;

Qy         58 GHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSC 117
              |: | : :  :  :|  | ||:     : ||| ||:| |   | |  | |:| |:  | |
Db         21 GNAASARHHGLLASA--RQPGVCHYGTKLACCYGWRRNS--KGVC-EATCEPGCK-FGEC 74

Qy        118 VQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCV 177
              | | :|||  |: | ||  ||:||  :   |  ||:|| ||| | |  || |  | | ||
Db         75 VGPNKCRCFPGYTGKTCSQDVNECGMKPRPCQHRCVNTHGSYKCFCLSGHMLMPDAT-CV 133

Qy        178 ----------------PKGGPP--------RVAPN 188
                               : ||         |:|||
Db        134 NSRTCAMINCQYSCEDTEEGPQCLCPSSGLRLAPN 168


RESULT 14
Q8IUX8
ID   Q8IUX8      PRELIMINARY;      PRT;   553 AA.
AC   Q8IUX8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to EGF-like-domain, multiple 6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC038587; AAH38587.1; -.
```

```
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA   Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA   Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY061220; AAL28768.1; -.
DR   EMBL; AE003481; AAF47894.2; -.
DR   FlyBase; FBgn0035539; CG7447.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   EGF-like domain.
```

```
SQ   SEQUENCE   512 AA;  57770 MW;  4DD372E4A6FE7627 CRC64;

  Query Match             18.4%;  Score 277.5;  DB 5;  Length 512;
  Best Local Similarity   27.8%;  Pred. No. 1.3e-16;
  Matches   73;  Conservative   40;  Mismatches  113;  Indels   37;  Gaps    9;

Qy          28 RRVC-AVRAHGDPV--SESFVQRVYQPFLTTCD----GHRACSTYRTIYRTAYRRSPGLA 80
               | :|   |    ||  :| : :  ::   | |       : |:  : ::  |||
Db         245 RHICMQQRTVTMPVKRTEVYSRPTWKHVATPCQPPTFSGQKCTRVQVVHEQAYRDVIDHK 304

Qy          81 PARP-RYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVD 139
                |:   | || || | :    :|   ||   |:|||:|  |  | || |: |  |: |||
Db         305 TAQQMTYDCCTGWSRENPRSDSCMKPICSARCQNGGNCTAPSTCSCPTGFTGRFCEQDVD 364

Qy         140 ECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKGGPPRVAPN----------P 189
               ||   :  | |:|||| |||:|:| :|  | :|   |       :|: |
Db         365 ECQTEK-PCDQQCINTHGSYFCRCRQGFVLQSDQQSC------KKVSTNADDAFEARDLE 417

Qy         190 TGVDSAMKEEVQRLQS----------RVDLLEEKLQLVLAPLHSLASQ--ALEHGLPDPG 237
                 :|    |   |||             : |:: ||   : : :| |:   ||    |
Db         418 NDIDDTDAEVATRLQKIEKSLANERVHTNELQKSLQATYSVVDTLKSRLSTLEKQAQDVS 477

Qy         238 SLLVHSFQQLGRIDSLSEQISFL 260
                |  : ::   | : |   :: |
Db         478 RLQTNLYKTESRTNKLEGMLNLL 500


RESULT 11
O88281
ID   O88281     PRELIMINARY;      PRT;  1574 AA.
AC   O88281;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MEGF6.
GN   MEGF6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RX   MEDLINE=98360089; PubMed=9693030;
RA   Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT   "Identification of high-molecular-weight proteins with multiple EGF-
RT   like motifs by motif-trap screening.";
RL   Genomics 51:27-34(1998).
DR   EMBL; AB011532; BAA32462.1; -.
DR   PIR; T13954; T13954.
DR   HSSP; P00736; 1APQ.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   Pfam; PF00008; EGF; 20.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00179; EGF_CA; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 5.
DR   PROSITE; PS00022; EGF_1; 23.
DR   PROSITE; PS01186; EGF_2; 23.
DR   PROSITE; PS01187; EGF_CA; 5.
KW   EGF-like domain.
SQ   SEQUENCE   1574 AA;  165445 MW;  2B48533D8F77F6E7 CRC64;

  Query Match             17.0%;  Score 256;  DB 11;  Length 1574;
  Best Local Similarity   37.9%;  Pred. No. 3.9e-14;
  Matches   64;  Conservative   16;  Mismatches   61;  Indels   28;  Gaps    9;

Qy          30 VCAVR-----AHGDPVSESFVQRVYQPFLTTCDGHRAC--STYRTIYRTAYRRSPGLAPA 82
               ||| :      |  |  ::| : |     | |     |     ||:|  :||:      |
```

```
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U89336; AAB47494.1; -.
DR   EMBL; BC052591; AAH52591.1; -.
DR   HSSP; P35555; 1EMN.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   EGF-like domain.
SQ   SEQUENCE   293 AA;  32262 MW;  6519CA254FB23FD0 CRC64;

  Query Match             28.1%;  Score 423.5;  DB 4;  Length 293;
  Best Local Similarity   36.5%;  Pred. No. 5e-30;
  Matches  103;  Conservative   37;  Mismatches  107;  Indels   35;  Gaps   10;

Qy         12 LLVLAVGGTEHAYRPGRRVCAVRAHGDPV--SESFVQRVYQPFLTTCDGHRACSTYRTIY 69
              ||:   |    : |  : ||: :    |:  :||: | ||:|:|| | | | ||||||:|
Db         19 LLIPGEGAKGGSLRESQGVCSKQTLVVPLHYNESYSQPVYKPYLTLCAGRRICSTYRTMY 78

Qy         70 RTAYRRSPGLAPARPRYACCPGWKRTSGLPGA--CGAAICQPPCRNGGSCVQPGRCRCPA 127
              |  : |       :    || |||:    |||  |  |||  || ||| ||:| :| |
Db         79 RVMW-REVRREVQQTHAVCCQGWKKRH--PGALTC-EAICAKPCLNGGVCVRPDQCECAP 134

Qy        128 GWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKGGPPRVAP 187
              || |  |  |||||      |   | |||||: | |     |  ||  |:     |  :
Db        135 GWGGKHCHVDVDECRTSITLCSHHCFNTAGSFTCGCPHDLVLGVDGRTCMEGSPEPPTSA 194

Qy        188 NPTGV--------DSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQA---LEHGLP-D 235
              :   |        : |:|:|:  |: |::  ||:            | ||   :   ||
Db        195 SILSVAVREAEKDERALKQEIHELRGRLERLEQ-----------WAGQAGAWVRAVLPVP 243

Qy        236 PGSLLVHSFQQL----GRIDSLSEQISFLEEQLGSCSCKKDS 273
              |  |      :|     ||:|||:|:  |||:||:|||: :|
Db        244 PEELQPEQVAELWGRGDRIESLSDQVLLLEERLGACSCEDNS 285


RESULT 9
Q8IV30
ID   Q8IV30      PRELIMINARY;      PRT;   293 AA.
AC   Q8IV30;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Chromosome 6 open reading frame 8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC035574; AAH35574.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   293 AA;  32261 MW;  6519CA255568FFD0 CRC64;

  Query Match             27.9%;  Score 420.5;  DB 4;  Length 293;
  Best Local Similarity   36.2%;  Pred. No. 9.4e-30;
  Matches  102;  Conservative   38;  Mismatches  107;  Indels   35;  Gaps   10;

Qy         12 LLVLAVGGTEHAYRPGRRVCAVRAHGDPV--SESFVQRVYQPFLTTCDGHRACSTYRTIY 69
              ||:   |    : |  : ||: :    |:  :||: | ||:|:|| | | | ||||||:|
Db         19 LLIPGEGAKGGSLRESQGVCSKQTLVVPLHYNESYSQPVYKPYLTLCAGRRICSTYRTMY 78

Qy         70 RTAYRRSPGLAPARPRYACCPGWKRTSGLPGA--CGAAICQPPCRNGGSCVQPGRCRCPA 127
              |  : |       :    || |||:    |||  |  |||  || ||| ||:| :| |
Db         79 RVMW-REVRREVQQTHAVCCQGWKKRH--PGALTC-EAICAKPCLNGGVCVRPDQCECAP 134

Qy        128 GWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKGGPPRVAP 187
              || |  |  |||||      |   | |||||: | |     |  ||  |:     |  :
Db        135 GWGGKHCHVDVDECRTSITLCSHHCFNTAGSFTCGCPHDLVLGVDGRTCMEGSPEPPTSA 194

Qy        188 NPTGV--------DSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQA---LEHGLP-D 235
              :   |        : |:|:|:  |: |::  ||:            | ||   :   ||
Db        195 SILSVAVREAEKDERALKQEIHELRGRLERLEQ-----------WAGQAGAWVRAVLPVP 243

Qy        236 PGSLLVHSFQQL----GRIDSLSEQISFLEEQLGSCSCKKDS 273
              |  |      :|     ||:|||:|:  |:|:||:|||: :|
Db        244 PEELQPEQVAELWGRGDRIESLSDQVLLLQERLGACSCEDNS 285


RESULT 10
Q95RQ1
ID   Q95RQ1      PRELIMINARY;      PRT;   512 AA.
AC   Q95RQ1; Q9VZD0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   LD16414p (CG7447-PA) (CG7447-PB).
GN   CG7447.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
Db        141 LASRSPEHGLQDPGSLLAHSFQQLDRIDSLSEQVSFLEEQLGSCSCKKD 189

RESULT 6
Q7ZXT0
ID   Q7ZXT0      PRELIMINARY;      PRT;   280 AA.
AC   Q7ZXT0;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to NEU1 protein.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC044267; AAH44267.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
SQ   SEQUENCE   280 AA;  30720 MW;  424EF8CF1B16D220 CRC64;

  Query Match             45.3%;  Score 682.5;  DB 13;  Length 280;
  Best Local Similarity   48.7%;  Pred. No. 2.2e-53;
  Matches  132;  Conservative   44;  Mismatches   84;  Indels   11;  Gaps    6;

Qy         11 WLLVLAV--GGTEHAYRPGRRVCAVRAHGDPVS--ESFVQRVYQPFLTTCDGHRACSTYR 66
              :||:|||     :| || |||:|:   |   ||  :|||| |: | :| |:||| |||||
Db         11 YLLILAVTSAAADHLYRTGRRICSADGHPGTVSVTQSFVQPVHSPIMTLCEGHRICSTYR 70

Qy         67 TIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCRCP 126
              | |: :||:      : | |:|||||:|      :|| |:|:  |:|||:||   :| ||
Db         71 TTYKVSYRQ-VSRKTSFPLYSCCPGWRRIGAQTHSCGQALCRLQCQNGGTCVSSNKCECP 129

Qy        127 AGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLC--VPKGGPPR 184
              |||||  || ||||||     | | |||:|||: |:| ||: | |||  |  ||    |
Db        130 AGWRGIHCQMDVDECSDGTHQCSQACINSAGSFSCECLEGYRLMADGKTCRKVPAPTVPP 189

Qy        185 VAP---NPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLLV 241
               :|     :|:  ::|||:  |:|::|:||:|| |:| |   | : : :    || :||
Db        190 ASPTSVQESGIPHSVKEEMAELRSKIDVLEQKLHLLLTPFQGLTTFSPDDA-ADPIALLT 248

Qy        242 HSFQQLGRIDSLSEQISFLEEQLGSCSCKKD 272
               | ||| ||||||||||||||:| :|||| :
Db        249 RSLQQLDRIDSLSEQISFLEERLETCSCKTE 279

RESULT 7
O35447
ID   O35447      PRELIMINARY;      PRT;   293 AA.
AC   O35447;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA   Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT   "Sequence of the mouse major histocompatibility locus class III
RT   region.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF030001; AAB82010.1; -.
DR   PIR; T09065; T09065.
DR   HSSP; P35555; 1EMN.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   293 AA;  32066 MW;  8A99A5EC000A2C62 CRC64;

  Query Match             28.9%;  Score 435.5;  DB 11;  Length 293;
  Best Local Similarity   37.3%;  Pred. No. 4.2e-31;
  Matches  112;  Conservative   36;  Mismatches  101;  Indels   51;  Gaps   14;

Qy          1 MRGSQEVLLMWLLVLAVG-GTE-HAYRPGRRVCAVRAHGDPV--SESFVQRVYQPFLTTC 56
              :||     | : |||  | ||   :::    ||: :    |:  :||: | ||:|:|| |
Db         11 LRG-----LSFFLVLMTGEGTRGGSFKESLGVCSKQTLLVPLRYNESYSQPVYKPYLTLC 65

Qy         57 DGHRACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGA-AICQPPCRNGG 115
               | | |||||| || |:|      | :    || |||:    |||    |||  || |||
Db         66 AGRRICSTYRTTYRVAWREVRREVP-QTHVVCCQGWKKPH--PGALTCDAICSKPCLNGG 122

Qy        116 SCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTL 175
               |  | || |  || |  |  ||||| |    |   |:|| ||: | |     |  ||
Db        123 VCTGPDRCECAPGWGGKHCHVDVDECRASLTLCSHGCLNTLGSFLCSCPHPLVLGLDGRT 182

Qy        176 CVPKGGPPRVAPNPTGV--------------DSAMKEEVQRLQSRVDLLEEKLQLVLAPL 221
              |   ||||    :||                : |:: ||  |: |:: ||:
Db        183 CA--GGPPE---SPTSASILSVAVREADSEEERALRWEVAELRGRLEKLEQ--------- 228

Qy        222 HSLASQA---LEHGLP-DPGSLLVHSFQQL----GRIDSLSEQISFLEEQLGSCSCKKDS 273
                 |:||   :   ||  |  |      :|     ||:|||:|:  |||:||:|:|: :|
Db        229 --WATQAGAWVRAVLPMPPEELRPEQVAELWGRGDRIESLSDQVLLLEERLGACACEDNS 286

RESULT 8
Q99944
ID   Q99944      PRELIMINARY;      PRT;   293 AA.
AC   Q99944;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NG3 (C6orf8 protein).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA   Banta A., Spies T., Hood L.;
RT   "Sequence determination of 300 kilobases of the human class III MHC
RT   locus.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

```
                 |||||||||||||||||||: |||||||||||||||||||||||||||||||| |||||::
Db          64 RACSTYRTIYRTAYRRSPGVTPARPRYACCPGWKRTSGLPGACGAAICQPPCGNGGSCIR 123

Qy         120 PGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPK 179
               || |||| ||:|||||:||||||     |||||:|| |||||| ||| | ||||| |: |
Db         124 PGHCRCPVGWQGDTCQTDVDECSTGEASCPQRCVNTVGSYWCQGWEGQSPSADGTRCLSK 183

Qy         180 GGPPRVAPNPT-GVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGS 238
                ||  |||||| ||||  :||| |||:|||:||:||||||||||||||:: |||| ||||
Db         184 EGPSPVAPNPTAGVDSMAREEVYRLQARVDVLEQKLQLVLAPLHSLASRSTEHGLQDPGS 243

Qy         239 LLVHSFQQLGRIDSLSEQISFLEEQLGSCSCKKD 272
               || |||||| ||||||||:||||| |||||||||
Db         244 LLAHSFQQLDRIDSLSEQVSFLEEHLGSCSCKKD 277


RESULT 4
Q9DCP5
ID   Q9DCP5      PRELIMINARY;      PRT;   265 AA.
AC   Q9DCP5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Vascular endothelial zinc finger 1.
GN   VEZF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Kidney;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK002601; BAB22222.1; -.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:1313291; Vezf1.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   EGF-like domain.
SQ   SEQUENCE   265 AA;  28665 MW;  D81EC3DA884FF97E CRC64;

  Query Match             73.3%;  Score 1103.5;  DB 11;  Length 265;
  Best Local Similarity   73.7%;  Pred. No. 2.6e-91;
  Matches  202;  Conservative   18;  Mismatches   39;  Indels   15;  Gaps    3;

Qy           1 MRGSQEVLLMWLLVLAVGG-TEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGH 59
```

```
               | || |:|: | ||||  | ||| ||| |||| |   |  :||:||||||||:|||||||
Db           4 MWGSGELLVAWFLVLAADGTTEHVYRPSRRVCTVGISGGSISETFVQRVYQPYLTTCDGH 63

Qy          60 RACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQ 119
               |||||||||||||||||||: |||||||||||||||||||||||||||||||| |||||::
Db          64 RACSTYRTIYRTAYRRSPGVTPARPRYACCPGWKRTSGLPGACGAAICQPPCGNGGSCIR 123

Qy         120 PGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPK 179
               || |||| ||:|||||:||||||     |||||:|| |||||| ||| | ||||| |: |
Db         124 PGHCRCPVGWQGDTCQTDVDECSTGEASCPQRCVNTVGSYWCQGWEGQSPSADGTRCLSK 183

Qy         180 GGPPRVAPNPT-GVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGS 238
                ||  |||||| ||||  :||| |||:|||:||:||||||||||||||:: |||| ||||
Db         184 EGPSPVAPNPTAGVDSMAREEVYRLQARVDVLEQKLQLVLAPLHSLASRSTEHGLQDPGS 243

Qy         239 LLVHSFQQLGRIDSLSEQISFLEEQLGSCSCKKD 272
               || |||||| ||||||                |||||
Db         244 LLAHSFQQLDRIDSLS-------------SCKKD 264


RESULT 5
Q9JKW3
ID   Q9JKW3      PRELIMINARY;      PRT;   190 AA.
AC   Q9JKW3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CBL20.
GN   CBL20.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RA   Chan M.T.W., Ng C.C.Y., Lim E.K.B., Huynh H.T.;
RT   "Cloning and Characterization of a Novel 20.4kD Estrogen-regulated
RT   protein in the Rat Spleen.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RA   Marcantonio D., Huynh H.T.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF223678; AAF35352.1; -.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   EGF-like domain.
SQ   SEQUENCE   190 AA;  20527 MW;  C540EF0687F1E998 CRC64;

  Query Match             48.3%;  Score 727.5;  DB 11;  Length 190;
  Best Local Similarity   76.9%;  Pred. No. 1.3e-57;
  Matches  130;  Conservative   15;  Mismatches   23;  Indels    1;  Gaps    1;

Qy         105 AICQPPCRNGGSCVQPGRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCW 164
               ||||||| | |||::||||||| ||:||||| ||||||     |||||:|| ||||||||
Db          21 AICQPPCGNEGSCIRPGRCRCPVGWQGDTCQIDVDECSTGEARCPQRCVNTVGSYWCQCW 80

Qy         165 EGHSLSADGTLCVPKGGPPRVAPNPT-GVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHS 223
               || | |||| ||:|| ||  |||:|| |||| ::| | :||:|||:||:|||||||||||
Db          81 EGQSPSADGVLCLPKEGPSPVAPSPTPGVDSVVREXVYKLQARVDVLEQKLQLVLAPLHS 140

Qy         224 LASQALEHGLPDPGSLLVHSFQQLGRIDSLSEQISFLEEQLGSCSCKKD 272
               |||:: |||| |||||| |||||| ||||||||:|||||||||||||||
```

```
Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180

Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 2
Q9UHF1
ID   Q9UHF1      PRELIMINARY;      PRT;   273 AA.
AC   Q9UHF1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NOTCH4-like protein (Hypothetical protein).
GN   ZNEU1 OR DKFZP586L2317.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sheppard P., Jelinek L., Whitmore T., Blumberg H., Lehner J.,
RA   O'Hara P.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Koehrer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF186111; AAF01429.1; -.
DR   EMBL; AL512735; CAC21666.1; -.
DR   HSSP; P08709; 1FAK.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   273 AA;  29617 MW;  5740BB845ED5A988 CRC64;

  Query Match             99.9%;  Score 1504;  DB 4;  Length 273;
  Best Local Similarity   99.6%;  Pred. No. 2.2e-127;
  Matches  272;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60

Qy         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQP 120

Qy        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKG 180
              ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db        121 GRCRCPAGWRGDTCQSDVDECSARRGGCPQRCVNTAGSYWCQCWEGHSLSADGTLCVPKG 180
```

```
Qy        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GPPRVAPNPTGVDSAMKEEVQRLQSRVDLLEEKLQLVLAPLHSLASQALEHGLPDPGSLL 240

Qy        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273
              |||||||||||||||||||||||||||||||||
Db        241 VHSFQQLGRIDSLSEQISFLEEQLGSCSCKKDS 273


RESULT 3
Q9QXT5
ID   Q9QXT5      PRELIMINARY;      PRT;   278 AA.
AC   Q9QXT5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   NOTCH4-like protein (Vascular endothelial zinc finger 1) (EGF-like
DE   domain 7).
GN   EGFL7 OR VEZF1 OR ZNEU1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sheppard P., Jelinek L., Whitmore T., Blumberg H., Lehner J.,
RA   O'Hara P.;
RT   "Mus musculus homolog of HLA class III region containing NOTCH4
RT   gene.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Presnell S., Gilbert T., Whitmore T., Foster D., Hart C., Lehner J.,
RA   Martinez T., Hoffman R., O'Hara P.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD-1;
RA   Fitch M.J., Kuhnert F., Stuhlmann H.;
RT   "Isolation of an early murine embryonic gene, Egfl7.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF184973; AAF01322.1; -.
DR   EMBL; BC024610; AAH24610.1; -.
DR   EMBL; AY309459; AAP74732.1; -.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:1313291; Vezf1.
DR   MGD; MGI:2449923; Egfl7.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   EGF-like domain.
SQ   SEQUENCE   278 AA;  30125 MW;  863159F567102FA7 CRC64;

  Query Match             78.4%;  Score 1180;  DB 11;  Length 278;
  Best Local Similarity   77.7%;  Pred. No. 3.5e-98;
  Matches  213;  Conservative   19;  Mismatches   40;  Indels    2;  Gaps    2;

Qy          1 MRGSQEVLLMWLLVLAVGG-TEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGH 59
              | || |:|: | |||| |  ||| ||| |||| |   |  :||:||||||||:|||||||
Db          4 MWGSGELLVAWFLVLAADGTTEHVYRPSRRVCTVGISGGSISETFVQRVYQPYLTTCDGH 63

Qy         60 RACSTYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQ 119
```

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         April 22, 2004, 12:42:40 ; Search time 45 Seconds
                                            (without alignments)
                                            1914.144 Million cell updates/sec

Title:          US-09-978-191A-506
Perfect score:  1505
Sequence:       1 MRGSQEVLLMWLLVLAVGGT..........SEQISFLEEQLGSCSCKKDS 273

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters:      1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_25:*
               1:  sp_archea:*
               2:  sp_bacteria:*
               3:  sp_fungi:*
               4:  sp_human:*
               5:  sp_invertebrate:*
               6:  sp_mammal:*
               7:  sp_mhc:*
               8:  sp_organelle:*
               9:  sp_phage:*
               10:  sp_plant:*
               11:  sp_rodent:*
               12:  sp_virus:*
               13:  sp_vertebrate:*
               14:  sp_unclassified:*
               15:  sp_rvirus:*
               16:  sp_bacteriap:*
               17:  sp_archeap:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result              Query
  No.      Score    Match Length DB  ID                        Description
-----------------------------------------------------------------------------------
    1      1505   100.0     273  4  Q96EG0                    Q96eg0 homo sapien
    2      1504    99.9     273  4  Q9UHF1                    Q9uhf1 homo sapien
    3      1180    78.4     278  11  Q9QXT5                    Q9qxt5 mus musculu
    4    1103.5    73.3     265  11  Q9DCP5                    Q9dcp5 mus musculu
    5     727.5    48.3     190  11  Q9JKW3                    Q9jkw3 rattus norv
    6     682.5    45.3     280  13  Q7ZXT0                    Q7zxt0 xenopus lae
    7     435.5    28.9     293  11  O35447                    O35447 mus musculu
    8     423.5    28.1     293  4  Q99944                    Q99944 homo sapien
    9     420.5    27.9     293  4  Q8IV30                    Q8iv30 homo sapien
   10     277.5    18.4     512  5  Q95RQ1                    Q95rq1 drosophila
   11       256    17.0    1574  11  O88281                    O88281 rattus norv
   12     252.5    16.8     558  4  Q9UFK6                    Q9ufk6 homo sapien
   13     251.5    16.7     553  4  Q9NZL7                    Q9nzl7 homo sapien
   14     251.5    16.7     553  4  Q8IUX8                    Q8iux8 homo sapien
   15     251.5    16.7     554  4  Q9NY67                    Q9ny67 homo sapien
   16     249.5    16.6     553  4  Q8NBV0                    Q8nbv0 homo sapien
```

```
   17     245.5    16.3     327  11  Q8BPM8                    Q8bpm8 mus musculu
   18     245.5    16.3     550  11  Q9JJZ5                    Q9jjz5 mus musculu
   19     225.5    15.0     544  13  Q8AVH7                    Q8avh7 xenopus lae
   20       223    14.8     561  11  Q91V88                    Q91v88 mus musculu
   21     218.5    14.5     578  11  Q91ZD3                    Q91zd3 mus musculu
   22       218    14.5    1664  5  Q9TVQ2                   Q9tvq2 caenorhabdi
   23     217.5    14.5     528  11  Q9CXD8                    Q9cxd8 mus musculu
   24       198    13.2     747  11  Q8VHF4                    Q8vhf4 mus musculu
   25       198    13.2    1004  11  Q8CGA7                    Q8cga7 mus musculu
   26       198    13.2    1034  11  Q8VHL7                    Q8vhl7 mus musculu
   27     197.5    13.1     592  11  Q91XL5                    Q91xl5 mus musculu
   28       196    13.0    1034  11  Q8VIK5                    Q8vik5 mus musculu
   29     194.5    12.9     590  11  Q8C088                    Q8c088 mus musculu
   30       193    12.8     609  11  Q923T5                    Q923t5 mus musculu
   31     191.5    12.7     814  13  Q7ZXL5                    Q7zxl5 xenopus lae
   32       189    12.6     678  4  Q14393                   Q14393 homo sapien
   33     185.5    12.3    1600  11  Q8K4G0                    Q8k4g0 mus musculu
   34     185.5    12.3    1666  11  Q8K4G1                    Q8k4g1 mus musculu
   35     184.5    12.3    2906  11  Q9WUH9                    Q9wuh9 rattus norv
   36       184    12.2     673  11  Q61592                    Q61592 mus musculu
   37       184    12.2     674  11  Q99K57                    Q99k57 mus musculu
   38     183.5    12.2     648  13  Q7T3H4                    Q7t3h4 brachydanio
   39     183.5    12.2     784  11  Q8BM43                    Q8bm43 mus musculu
   40     183.5    12.2     816  11  Q8R417                    Q8r417 rattus norv
   41     183.5    12.2     819  11  Q80UM5                    Q80um5 mus musculu
   42     183.5    12.2     858  11  Q8BM06                    Q8bm06 mus musculu
   43     182.5    12.1     815  4  Q96JS2                   Q96js2 homo sapien
   44     181.5    12.1     648  5  Q9VJU4                   Q9vju4 drosophila
   45     181.5    12.1     648  5  Q9NKD7                   Q9nkd7 drosophila

                             ALIGNMENTS

RESULT 1
Q96EG0
ID   Q96EG0      PRELIMINARY;      PRT;   273 AA.
AC   Q96EG0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to NEU1 protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC012377; AAH12377.1; -.
DR   Genew; HGNC:20594; EGFL7.
DR   GO; GO:0005509; F:calcium ion binding; IEA.
DR   InterPro; IPR000152; Asx_hydroxyl_S.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 2.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   EGF-like domain.
SQ   SEQUENCE   273 AA;  29631 MW;  5AD0A4845ED5B688 CRC64;

  Query Match             100.0%;  Score 1505;  DB 4;  Length 273;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-127;
  Matches  273;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHR 60
```